```
OM of: US-09-303-518D-463 to: SwissProt_40:*   out_format : pfs

Date: Jun 30, 2002  8:36 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
  -MODEL=frame+_n2p.model -DEV=xlh
  -Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142714_4362/app_query.fasta_1.23501
  -DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
  -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
  -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
  -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
  -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
  -TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
  -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
  -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
  -USER=US09303518_@CGN1_1_440 -NCPU=6 -ICPU=3 -LONGLOG
  -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-463
Query length: 1887
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 217.960000


score_list:
Sequence                  Strd Orig    ZScore    EScore Len   ! Documentation  ..
SwissProt_40:MUC2_HUMAN +  165.50  160.64  0.0058    5179  ! Q02817 homo sapiens (human)
SwissProt_40:FYB_MOUSE +  160.50  171.44  0.0091    819  ! O35601 mus musculus (mouse).
SwissProt_40:DAN4_YEAST +  153.50  160.49  0.0262    1161  ! P47179 saccharomyces cerevi
SwissProt_40:AMYH_YEAST +  146.00  150.64  0.0788    1367  ! P08640 saccharomyces cerevi
SwissProt_40:NKCR_MOUSE +  143.50  147.29  0.1139    1453  ! P30415 mus musculus (mouse)
SwissProt_40:YAV1_SCHPO +  143.50  145.42  0.1173    1794  ! Q10172 schizosaccharomyces
SwissProt_40:KP58_DROME +  141.00  148.25  0.1537    952  ! Q9vpc0 drosophila melanogast
SwissProt_40:YGY3_HALSQ +  139.00  152.94  0.1835    437  ! P21561 haloferax sp. (strain
SwissProt_40:DAB_DROME +  138.50  137.18  0.2510    2411  ! P98081 drosophila melanogast
SwissProt_40:PODX_RABIT +  138.00  149.76  0.2190    551  ! Q28645 oryctolagus cuniculus
SwissProt_40:MAP4_MOUSE +  138.00  143.40  0.2422    1125  ! P27546 mus musculus (mouse)
SwissProt_40:YWR1_CAEEL +  137.50  142.10  0.2633    1223  ! Q10925 caenorhabditis elega
SwissProt_40:VRP1_YEAST +  137.00  145.13  0.2673    817  ! P37370 saccharomyces cerevis
SwissProt_40:VGLX_HSVEB +  136.50  144.79  0.2862    797  ! P28968 equine herpesvirus ty
SwissProt_40:AMYB_PAEPO +  133.00  137.26  0.5012    1196  ! P21543 paenibacillus polymy
SwissProt_40:DRPL_HUMAN +  132.00  136.22  0.5779    1185  ! P54259 homo sapiens (human)
SwissProt_40:GTFB_STRMU +  129.50  131.46  0.8538    1476  ! P08987 streptococcus mutans
SwissProt_40:VGP3_EBVA8 +  129.00  135.44  0.8537    886  ! Q07284 epstein-barr virus (s
SwissProt_40:VE2_HPV08 +  128.50  140.01  0.8457    498  ! P06422 human papillomavirus t
SwissProt_40:DRPL_RAT +  126.50  130.07  1.27    1183  ! P54258 rattus norvegicus (rat
SwissProt_40:AGA1_YEAST +  125.00  132.74  1.47    725  ! P32323 saccharomyces cerevis
SwissProt_40:HIG_DROME +  124.50  129.71  1.65    958  ! Q09101 drosophila melanogaste
SwissProt_40:YQU3_CAEEL +  124.50  127.33  1.71    1251  ! Q09550 caenorhabditis elega
SwissProt_40:V70K_TYMVC +  124.00  132.90  1.67    628  ! P28478 turnip yellow mosaic
SwissProt_40:YS8A_CAEEL +  124.00  130.79  1.73    796  ! Q09625 caenorhabditis elegan
SwissProt_40:VGP3_EBV +  123.50  129.07  1.89    907  ! P03200 epstein-barr virus (str
SwissProt_40:EGR1_BRARE +  123.00  133.62  1.87    511  ! P26632 brachydanio rerio (ze
SwissProt_40:A180_MOUSE +  123.00  128.57  2.03    901  ! Q61548 mus musculus (mouse).
SwissProt_40:MGPC_MYCGE +  123.00  127.19  2.07    1052  ! P22747 mycoplasma genitaliu
SwissProt_40:GTFC_STRMU +  123.00  124.81  2.15    1375  ! P13470 streptococcus mutans
SwissProt_40:SMF1_HUMAN +  123.00  121.92  2.25    1902  ! O14497 homo sapiens (human)
SwissProt_40:V70K_TYMVA +  122.50  131.22  2.07    628  ! P20131 turnip yellow mosaic
SwissProt_40:NCR1_HUMAN +  121.50  118.02  2.89    2440  ! O75376 homo sapiens (human)
SwissProt_40:PSC_DROME +  121.00  121.20  2.93    1603  ! P35820 drosophila melanogast
SwissProt_40:N214_HUMAN +  120.50  118.28  3.27    2090  ! P35658 homo sapiens (human)
SwissProt_40:ALP_CEPAC +  120.00  132.39  2.78    402  ! P29118 cephalosporium acremon
SwissProt_40:MUC4_HUMAN +  120.00  128.68  2.95    610  ! Q99102 homo sapiens (human).
SwissProt_40:DAG1_MOUSE +  119.50  124.73  3.35    893  ! Q62165 mus musculus (mouse).
SwissProt_40:CA1E_CHICK +  119.50  118.07  3.72    1888  ! P32018 gallus gallus (chick
SwissProt_40:P1P_LACLC +  119.50  118.00  3.73    1902  ! P16271 lactococcus lactis (s
SwissProt_40:CCAA_MOUSE +  119.50  116.85  3.79    2164  ! P97445 mus musculus (mou
SwissProt_40:ITA4_HUMAN -  119.00  122.83  3.67    1038  ! P13612 homo sapiens (hum
SwissProt_40:MAP4_BOVIN +  118.50  121.98  3.97    1072  ! P36225 bos taurus (bovin
SwissProt_40:P2P_LACLC +  118.50  116.88  4.30    1902  ! P15293 lactococcus lactis
SwissProt_40:CCAA_RAT +  118.50  115.54  4.39    2212  ! P54282 rattus norvegicus (
SwissProt_40:CDGT_BACSS +  118.00  124.99  4.03    718  ! P31747 bacillus sp. (stra
SwissProt_40:MAP4_HUMAN +  118.00  120.78  4.31    1152  ! P27816 homo sapiens (hum
SwissProt_40:A180_RAT +  117.50  122.27  4.48    915  ! Q05140 rattus norvegicus (r
SwissProt_40:LG1_MAIZE +  117.00  129.09  4.28    399  ! O04003 zea mays (maize). l
SwissProt_40:STUA_EMENI +  117.00  125.61  4.52    590  ! P36011 emericella nidulan
SwissProt_40:V70K_TYMV +  117.00  125.06  4.56    628  ! P10357 turnip yellow mosai
SwissProt_40:IGA0_HAEIN +  117.00  116.23  5.25    1694  ! P44969 haemophilus influ
SwissProt_40:IGA2_HAEIN +  117.00  116.19  5.25    1702  ! P45384 haemophilus influ
SwissProt_40:MI15_CAEEL +  116.50  119.67  5.29    1080  ! Q23356 caenorhabditis el
SwissProt_40:YM96_YEAST +  116.00  118.63  5.73    1140  ! Q04893 saccharomyces cer
SwissProt_40:IGA4_HAEIN +  116.00  114.33  6.14    1849  ! P45386 haemophilus influ
SwissProt_40:GTF2_STRDO +  115.50  115.10  6.46    1592  ! P27470 streptococcus dow
SwissProt_40:P3P_LACLC +  114.50  112.40  7.64    1902  ! P15292 lactococcus lactis
SwissProt_40:YHC8_YEAST +  114.00  122.03  6.99    605  ! P38739 saccharomyces cere
SwissProt_40:EGR1_HUMAN +  113.50  122.43  7.39    543  ! P18146 homo sapiens (huma
SwissProt_40:MUC1_XENLA +  113.50  120.67  7.60    662  ! Q05049 xenopus laevis (af
SwissProt_40:GBF_DICDI +  113.50  120.07  7.68    708  ! P36417 dictyostelium disco
SwissProt_40:YQ36_CAEEL +  113.50  117.33  8.02    963  ! Q09457 caenorhabditis ele
SwissProt_40:YJ9F_YEAST +  113.50  113.74  8.49    1442  ! P47169 saccharomyces cer
SwissProt_40:ZEP1_MOUSE +  113.50  108.20  9.27    2688  ! Q03172 mus musculus (mou
SwissProt_40:ICP0_HSVBK +  113.00  119.92  8.19    676  ! P29836 bovine herpesvirus
SwissProt_40:SON_MOUSE +  113.00  108.63  9.80    2404  ! Q9qx47 mus musculus (mous
SwissProt_40:MUC1_MESAU +  112.50  119.36  8.80    676  ! Q60528 mesocricetus aurat
SwissProt_40:SRP_DROME +  112.50  118.10  8.98    779  ! P52172 drosophila melanoga
SwissProt_40:FLO5_YEAST +  112.50  115.23  9.40    1075  ! P38894 saccharomyces cer
SwissProt_40:TCOF_HUMAN +  112.50  112.81  9.77    1411  ! Q13428 homo sapiens (hum
SwissProt_40:FLO1_YEAST +  112.50  112.05  9.89    1537  ! P32768 saccharomyces cer
SwissProt_40:NCR2_HUMAN +  112.50  107.66  10.60    2517  ! Q9y618 h nuclear recepto
SwissProt_40:ICP0_HSVBK -  112.00  118.80  9.46    676  ! P29836 bovine herpesvirus
SwissProt_40:YMJB_CAEEL +  112.00  116.08  9.88    918  ! P34487 caenorhabditis ele
SwissProt_40:NIA_SPIOL +  112.00  116.00  9.89    926  ! P23312 spinacia oleracea (
SwissProt_40:PGCA_RAT +  112.00  108.61  11.12    2124  ! P07897 rattus norvegicus (
SwissProt_40:SON_HUMAN +  112.00  107.43  11.33    2426  ! P18583 homo sapiens (huma
SwissProt_40:WETA_PENCH +  111.50  120.94  9.74    499  ! Q01870 penicillium chryso
SwissProt_40:VG37_BPAR1 +  111.50  113.88  10.89    1103  ! Q9g0b5 bacteriophage ar1
SwissProt_40:IGA_NEIGO +  111.50  110.96  11.41    1532  ! P09790 neisseria gonorrho
SwissProt_40:YEW2_YEAST +  111.50  109.76  11.63    1753  ! P32634 saccharomyces cer
SwissProt_40:NCA1_XENLA +  111.00  113.44  11.68    1088  ! P16170 xenopus laevis (a
SwissProt_40:FURC_DROME +  111.00  113.34  11.70    1101  ! P30430 drosophila melano
SwissProt_40:YAG3_YEAST +  111.00  111.71  12.01    1322  ! P39712 saccharomyces cer
SwissProt_40:TUSP_MOUSE +  111.00  110.31  12.28    1547  ! Q9ji15 mus musculus (mou
SwissProt_40:FIG2_YEAST +  111.00  109.96  12.34    1609  ! P25653 saccharomyces cer
SwissProt_40:TENS_CHICK +  111.00  109.25  12.49    1744  ! Q04205 gallus gallus (ch
SwissProt_40:TRHY_HUMAN +  111.00  108.49  12.64    1898  ! Q07283 homo sapiens (hum
SwissProt_40:YMH7_CAEEL +  110.50  122.24  10.82    380  ! P34474 caenorhabditis ele
SwissProt_40:MANA_RHOMR +  110.50  113.45  12.44    1021  ! P49425 rhodothermus mari
SwissProt_40:SED1_YEAST +  110.00  122.72  11.44    338  ! Q01589 saccharomyces cere
SwissProt_40:LEUK_HUMAN +  110.00  121.23  11.71    400  ! P16150 homo sapiens (huma
SwissProt_40:CYP8_CAEEL +  110.00  119.87  11.97    466  ! P52016 caenorhabditis ele
SwissProt_40:ITF2_MOUSE +  110.00  116.64  12.60    670  ! Q60722 mus musculus (mous
SwissProt_40:CLPB_MYCTU +  110.00  114.54  13.02    848  ! O53719 mycobacterium tube
SwissProt_40:NKCR_HUMAN +  110.00  109.69  14.06    1462  ! P30414 homo sapiens (hum
SwissProt_40:TEGU_HSVEB +  110.00  102.13  15.85    3421  ! P28955 equine herpesviru
SwissProt_40:YK82_MYCTU +  109.50  115.42  13.67    721  ! Q10690 mycobacterium tube
SwissProt_40:PBPA_BACSU +  109.50  113.31  14.14    914  ! P39793 bacillus subtilis.

seq_name: SwissProt_40:MUC2_HUMAN

seq_documentation_block:
ID   MUC2_HUMAN     STANDARD;      PRT;   5179 AA.
AC   Q02817; Q14878;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 2 precursor (Intestinal mucin 2).
GN   MUC2 OR SMUC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=94132002; PubMed=8300571;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT   "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT   Identification of the amino terminus and overall sequence similarity
RT   to prepro-von Willebrand factor.";
RL   J. Biol. Chem. 269:2440-2446(1994).
RN   [2]
RP   SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=93016075; PubMed=1400449;
RA   Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA   Kim Y.S.;
RT   "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT   both upstream and downstream of its central repetitive region.";
RL   J. Biol. Chem. 267:21375-21383(1992).
RN   [3]
RP   SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX   MEDLINE=91358717; PubMed=1885763;
RA   Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA   Petersen G.M., Kim Y.S.;
RT   "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT   and polymorphism.";
RL   J. Clin. Invest. 88:1005-1013(1991).
CC   -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC       OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC       PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC       AGENTS AT MUCOSAL SURFACES.
CC   -!- SUBUNIT: MULTIMERIC.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC       BRONCHUS, CERVIX AND GALL BLADDER.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC       INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC   -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC       VARIES AMONG DIFFERENT ALLELES.
CC   -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC       OF SILKWORM HEMOCYTIN.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L21998; AAB95295.1; -.
DR   EMBL; M74027; AAA59875.1; -.
DR   EMBL; M94131; AAA59163.1; -.
DR   EMBL; M94132; AAA59164.1; -.
DR   MIM; 158370; -.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR001007; VWFC.
DR   InterPro; IPR001846; Vwd.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00094; vwd; 4.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   SMART; SM00214; VWC; 2.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   5179       MUCIN 2.
FT   DOMAIN     1401   1747       APPROXIMATE REPEATS.
FT   REPEAT     1401   1416       1.
FT   REPEAT     1417   1432       2.
FT   REPEAT     1433   1448       3.
FT   REPEAT     1449   1464       4.
FT   REPEAT     1465   1471       5.
FT   REPEAT     1472   1478       6.
FT   REPEAT     1479   1494       7A.
FT   REPEAT     1495   1517       7B.
FT   REPEAT     1518   1533       8A.
FT   REPEAT     1534   1556       8B.
FT   REPEAT     1557   1572       9A.
FT   REPEAT     1573   1596       9B.
FT   REPEAT     1597   1612       10A.
FT   REPEAT     1613   1635       10B.
FT   REPEAT     1636   1651       11A.
FT   REPEAT     1652   1675       11B.
FT   REPEAT     1676   1683       12.
FT   REPEAT     1684   1699       13.
FT   REPEAT     1700   1715       14.
FT   REPEAT     1716   1731       15.
FT   REPEAT     1732   1747       16.
FT   DOMAIN     4815   4886       VWFC 1.
FT   DOMAIN     4924   4991       VWFC 2.
FT   DOMAIN     5075   5160       CTCK.
FT   DISULFID   5075   5122       BY SIMILARITY.
FT   DISULFID   5089   5136       BY SIMILARITY.
FT   DISULFID   5098   5152       BY SIMILARITY.
FT   DISULFID   5102   5154       BY SIMILARITY.
FT   DISULFID      ?   5159       BY SIMILARITY.
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    670    670       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    770    770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1139   1139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1154   1154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1215   1215       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1230   1230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1246   1246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1787   1787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1820   1820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4339   4339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4351   4351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4362   4362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4373   4373       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4422   4422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4438   4438       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4502   4502       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4616   4616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4627   4627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4752   4752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4787   4787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4881   4881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4888   4888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4955   4955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4970   4970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5019   5019       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5038   5038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   5069   5069       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1351   1351       H -> L (IN REF. 3).
FT   CONFLICT   1412   1412       T -> S (IN REF. 3).
FT   CONFLICT   1449   1449       L -> P (IN REF. 3).
FT   CONFLICT   1504   1504       M -> T (IN REF. 3).
FT   CONFLICT   4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  85CD7571FB9A5663 CRC64;

alignment_scores:
               Quality:  165.50             Length:    334
                 Ratio:  1.226                Gaps:     11
```

```
Percent Similarity:  40.419   Percent Identity: 22.754

alignment_block:
 US-09-303-518D-463 x MUC2_HUMAN   ..

 Align seg 1/1  to: MUC2_HUMAN  from: 1  to: 5179

     269 ACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGAC 318
         |||   :::       |||   |||   |||:::::      |||   ||
    1520 ThrThrThrProSerProProThrThrThrThrThrThrProProProTh 1536

     319 AACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGG 368
         ||||   |||       |||       |||   |||:::::|||   |||:
    1536 rThrThrProSer...ProProThrThrThrProIleThrProProThrS 1552

     369 ATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCG 418
         :::::       |||:::         |||   |||   :::   ||||||
    1552 erThr......ThrThrLeuProProThrThrThrProSerProProPro 1566

     419 ACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCG 468
         |||:::   |||             ::::::   |||   |||
    1567 ThrThrThrThrThrProProProThrThrThrProSerPro........ 1580

     469 AGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCT 518
                     :::||||||      :::   |||   |||:::::::
    1581 ..........ProThrThrThrThrProSerProProThrIleThrThrT 1594

     519 CAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCC 568
         ::|||   |||   |||::::::|||:::             |||::::::
    1594 hrThrProProProThrThrThrProSerProProThrThrThrThrThr 1610

     569 ACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGC 618
         |||   |||                              |||   :::   ::
    1611 ThrProPro.......................ProThrThrThrProSe 1619

     619 GCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGC 668
         :||||||   |||:::|||               :::             |
    1619 rProProThrThrThrProIleThrProProThrSerThrThrThrLeuP 1636

     669 CTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAG 718
         ||   |||:::                           |||:::::::::
    1636 roProThrThrThrProSerProProProThrThrThrThrThrProPro 1652

     719 AAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAAC 768
               ::::::   :::   |||         :::         :::
    1653 ProThrThrThrProSerProProThrThrThrThrProSerProProIl 1669

     769 ATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGC 818
               :::   |||                     ||||||   |||
    1669 eThrThrThrThrThr..............ProProProThrThrThrP 1681

     819 GCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAG 868
            :::|||   |||                           |||   |||:::
    1681 roSerSerProIle....................ThrThrThrProSer 1690

     869 CAGCCA........................................TC 876
            |||                                          ||
    1691 ProProThrThrThrMetThrThrProSerProThrThrThrProSerSe 1707

     877 CGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGT 926
         |   |||       :::   ||||||               :::   :::|||
    1707 rProIleThrThrThrThrThrProSerSerThrThrThrProSerProP 1724

     927 CAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCC 976
                  ::::::      |||||||||:::
    1724 roProThrThrMetThrThrProSerProThr.................. 1734

     977 GGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCG 1026

    1734 .................................................. 1734

    1027 CAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAA 1076
                                             :::||||||::::::|||:

    1735 .....................ThrThrProSerProProThrThrThrM 1744

    1077 TTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATT...... 1120
         ::       :::       |||    ::::::    |||||||||
    1744 etThrThrLeuProProThrThrThrSerSerProLeuThrThrThrPro 1760

    1121 ......CCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAA 1164
               |||:::|||:::       |||:::|||       ::::::
    1761 LeuProProSerIleThrProProThrPheSerProPheSerThr..... 1775

    1165 AACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACT 1214
           |||:::|||       ||||||               :::   :::   ||||
    1776 .ThrThrProThrThrProCys...........ValProLeuCysAsnT 1788

    1215 GG 1216
          ||
    1788 rp 1788

seq_name: SwissProt_40:FYB_MOUSE

seq_documentation_block:
ID    FYB_MOUSE      STANDARD;      PRT;   819 AA.
AC    O35601; Q9Z2H3;
DT    15-JUL-1999 (Rel. 38, Created)
DT    16-OCT-2001 (Rel. 40, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    FYN-binding protein (FYN-T-binding protein) (FYB-120/130) (p120/p130)
DE    (SLP-76 associated phosphoprotein) (SLAP-130).
GN    FYB.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A. (ISOFORM FYB-120).
RC    TISSUE=T-cell lymphoma;
RX    MEDLINE=97352826; PubMed=9207119;
RA    da Silva A.J., Li Z., de Vera C., Canto E., Findell P., Rudd C.E.;
RT    "Cloning of a novel T-cell protein FYB that binds FYN and SH2-domain-
RT    containing leukocyte protein 76 and modulates interleukin 2
RT    production.";
RL    Proc. Natl. Acad. Sci. U.S.A. 94:7493-7498(1997).
RN    [2]
RP    SEQUENCE FROM N.A. (ISOFORM FYB-130).
RC    TISSUE=Hybridoma;
RX    MEDLINE=99428514; PubMed=10497204;
RA    Veale M., Raab M., Li Z., da Silva A.J., Kraeft S.-K., Weremowicz S.,
RA    Morton C.C., Rudd C.E.;
RT    "Novel isoform of lymphoid adaptor FYN-T-binding protein (FYB-130)
RT    interacts with SLP-76 and up-regulates interleukin 2 production.";
RL    J. Biol. Chem. 274:28427-28435(1999).
CC    -!- FUNCTION: ACTS AS A ADAPTER PROTEIN OF THE FYN AND SH2-DOMAIN-
CC        CONTAINING LEUCOCYTE PROTEIN-76 (SLP76) SIGNALING CASCADES IN T
CC        CELLS. MODULATES THE EXPRESSION OF INTERLEUKIN-2 (IL-2).
CC    -!- SUBUNIT: INTERACTS WITH FYN AND SLP76.
CC    -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC    -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; FYB-130 (SHOWN HERE) AND FYB-
CC        120; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    -!- TISSUE SPECIFICITY: EXPRESSED IN HEMATOPOIETIC TISSUES SUCH AS
CC        MYELOID AND T CELLS, SPLEEN AND THYMUS. NOT EXPRESSED IN B CELLS,
CC        NOR IN NON-LYMPHOID TISSUES. FYB-130 IS PREFERENTIALLY EXPRESSED
CC        IN MATURE T-CELLS COMPARED TO FYB-120, WHEREAS THYMOCYTES SHOWED A
CC        GREATER RELATIVE AMOUNT OF FYB-120.
CC    -!- PTM: T-CELL RECEPTOR LIGATION LEADS TO INCREASED TYROSINE
CC        PHOSPHORYLATION.
CC    -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC    the European Bioinformatics Institute.  There are no  restrictions on  its
CC    use  by  non-profit  institutions as long  as its content  is  in  no  way
CC    modified and this statement is not removed.  Usage  by  and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF001863; AAB62227.1; -.
DR   EMBL; AF061744; AAD03267.1; -.
DR   MGD; MGI:1346327; Fyb.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   SH3 domain; Phosphorylation; Nuclear protein; Coiled coil;
KW   Alternative splicing.
FT   DOMAIN      343    348       POLY-PRO.
FT   DOMAIN      381    385       POLY-PRO.
FT   DOMAIN      448    495       COILED COIL (POTENTIAL).
FT   DOMAIN      479    493       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      584    587       SH2-BINDING (TO SLP-76).
FT   DOMAIN      606    609       POLY-PRO.
FT   DOMAIN      615    618       SH2-BINDING (TO FYN).
FT   DOMAIN      622    625       POLY-ASP.
FT   DOMAIN      696    699       POLY-PRO.
FT   DOMAIN      699    756       SH3.
FT   DOMAIN      710    736       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      739    805       SH3-LIKE.
FT   VARSPLIC    627    672       MISSING (IN ISOFORM FYB-120).
SQ   SEQUENCE   819 AA;  90055 MW;  6222CC7EF1CA2BBD CRC64;

alignment_scores:
           Quality:  160.50            Length:    714
             Ratio:  0.544               Gaps:     31
 Percent Similarity:  41.317   Percent Identity: 20.308

alignment_block:
 US-09-303-518D-463 x FYB_MOUSE   ..

 Align seg 1/1  to: FYB_MOUSE  from: 1  to: 819

     125 TCGAACCCG..............ACGGGAAATACCACCTATTCGGCAGC 159
         ||||||||                |||:::::       :::::::::||
       8 SerAsnProThrGluGluAlaAlaThrSerSerArgProPheLysValAl 24

     160 AGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACA 209
         ||||   |||   |||||||:::
      24 aGlyGlnSerSerProSerGlyIleGln..................... 33

     210 AAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAA 259
                                   :::|||:::   |||    :::
      34 ...........................SerArgLysAsnLeuPheAsp 40

     260 ATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATT.. 307
           ::::::       :::|||  |||    :::|||   :::|||:::
      41 AsnGlnGly......AsnAlaSerProProAlaGlyProSerSerMetPr 55

     308 .CGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAG 356
          :::    :::||||||   |||     |||:::   :::||||||:::
      55 oLysPheGlyThrThrLysPro......ProLeuAlaAlaLysProThrT 70

     357 TCCCG...............................TTGACGGATTCA 373
                                             |||       :::
      70 yrGluGluLysProGluLysGluProLysProProPheLeuLysProThr 86

     374 GCCTTTACCGCATCCATTGGGACGGATACGAACACCATC........... 412
         :::   :::      :::||||||:::   |||::::::
      87 GlyGlySerProArgPheGlyThrGlnProAsnSerValSerArgAspPr 103

     413 ..................................CCGCCGACGGCTA 425
                                           ||||||:::
     103 oGluValLysValGlyPheLeuLysProValSerProLysProThrSerL 120

     426 TGACGGGCCACAGGGCGGCGGCTATCCCGCTC...CCAAAGGCGCGAGGG 472
         :::|||    :::    :::         :::   |||    |||   ::::::
     120 euThrLysGluAspSerLysProValValLeuArgProProGlyAsnLys 136
```

```
     473 ATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAAC 522
         ::::::::   :::        :::   |||          :::    :::
     137 LeuHisAsnLeuAsnGlnGluSerAspLeuLysThrProGlyProLysPr 153

     523 CTGACCGACAACCGCAGCACCGGACAACGGCTTG...CCGACCGTTTCCA 569
            |||:::      :::|||:::|||   |||   |||      |||
     153 oGlyProAlaProProValProGluAsnGluLeuLysProGlyPheSerL 170

     570 CA.......................ATGCCGGCGCTATGCTGACGCAAG 595
                                  |||||||||
     170 ysValAlaGlyAlaLysSerLysPheMetProAla.............. 181

     596 GAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGAC 645
               |||   |||:::              |||:::   |||   :::
     182 ......AlaGlnAspThr...........AspSerLysProArgPhePr 192

     646 AGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAA 695
         :::::::::         |||||||||   :::   :::   |||
     192 oArgHisThrPheGlyGlnLysProSerLeuSerThrGluAspSer.... 207

     696 AAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGC 745
                          ||||||:::   :::::::::   :::|||
     208 ................GlnGluGluAsnThrSerLysAsnValProVal 218

     746 AGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTG 795
         :::
     219 Gln.............................................. 219

     796 CTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGC 845
                                   ::::::|||   :::   |||:::
     220 .....................LysGlySerProVal...GlnLeuGlyA 228

     846 GCAACTCAAAGACTATGCCGCAGCAGCCA.................... 874
           :::||||||      |||   :::|||
     228 laLysSerLysGlyAlaProPheLysProProLysGluAspProGluAsp 244

     875 TCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCC 924
         :::      |||      ::::::|||:::|||         :::   |||||
     245 LysAspHisGlyAlaProSerSerProPheProGlyValValLeuLysPr 261

     925 GTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGT 974
         |:::|||   |||        :::|||:::   |||||::::::|||     :
     261 oAlaAla...SerArgGlySerProGlyLeuSerLysAsnPheGluGluL 277

     975 CCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATC............ 1012
         :::   |||:::                    :::::::::|||
     277 ysLysGluAspArgLysThrAspLeuAlaLysAsnIlePheLeuAsnLys 293

    1013 ..........................................CTGTC 1017
                                                   |||::
     294 LeuAsnGlnGluGluProAlaArgPheProLysAlaProSerLysLeuTh 310

    1018 AAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAAT....... 1060
         :::|||      |||:::::::|||:::    ::::::|||:::
     310 rAlaGlyThrProTrpGlyGlnSerGlnGluLysGluGlyAspLysAsnS 327

    1061 .........................CCGCCGTCAGCGACAATTTTGCCG 1084
                                  ||||||    :::::::::   |||
     327 erAlaThrProLysGlnLysAlaLeuProProLeuSerValLeuGlyPro 343

    1085 ATGCGGCATACGCCAAAT.....................ACCCGTCCCCT 1113
                     ||||||                       ||||||
     344 ProProProLysProAsnArgProProAsnValAspLeuThrArgPheAr 360

    1114 TACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGA 1163
                     :::   :::         |||   |||::::::         :::
     360 gLysAlaAspSerAlaAsnSerAlaThrLysSerGlnThrProTyrSerT 377

    1164 AAACATCACCTCCTCAACCGTGCCGCCGTCAAAC................ 1197
            ||||||   |||   |||   ||||||::::::
     377 hrThrSerLeuProProPro..ProProThrHisProAlaSerGlnProP 393

    1198 ..........................................GGC 1200
```

```
 393 roLeuProAlaSerHisProAlaHisProProValProSerLeuProPro 409

1201 AAAAATGTCAAA...CTGGCAGACCAACGCCACCCG............. 1233
     :::|||:::|||          |||    :::||||||
 410 ArgAsnIleLysProProLeuAspLeuLysHisProIleAsnAspGluAs 426

1234 .AAGACAGGCGTACCGTTT...GACGGTAAAGGGTTTCCGAATTTTGAGA 1279
       :::    ||||||        ||||||   |||      :::   |||:
 426 nGlnAspGlyValMetHisSerAspGlyThrGlyAsnLeuGluGluGluG 443

1280 AGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGC 1329
     :::::    :::    :::|||        ||||||:::::    :::
 443 lnGluSerGluGlyGluThrTyrGluAspIleAspSerSerLysGluArg 459

1330 GGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGT 1379
              :::   |||          :::   |||   |||   |||::
 460 AspLysLysArgGluLys.........GluGluLysLysArgLeuGluLe 473

1380 TGATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATG 1429
     :::||||||    :::|||          ||||||:::    |||::::::
 473 uGluArgLysGluGlnLys......GluArgGluLysLysGluGlnGluL 488

1430 TTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCG 1479
     ::::::::          :::   :::             |||     ||||||
 488 euLysLysLysPheLysLeuThrGlyProIleGlnValIleHisHisAla 504

1480 CAACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCATTTTATAGG 1529
     :::         :::   |||   |||   :::          |||
 505 LysAlaCysCysAspValLysGlyGlyLysAsnGluLeuSerPheLysGl 521

1530 TGGTGATATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCC 1579
       |||:::
 521 nGlyGlu.......................................... 523

1580 GTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGG 1629
            |||::::::::|||:::    |||:::
 524 .....AspIleGluIleIleArgIleThrAsp.................. 532

1630 GTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGT 1679
                                 :::|||:::|||:::|||
 533 ..........................AsnProGluGlyLysTrpLeuGl 540

1680 GAAAACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAA 1729
      :::|||    :::|||:::                |||   |||
 540 yArgThrAlaArgGlySerTyrGlyTyrIleLysThrThrAlaValGluI 557

1730 AAGATTGGGATGAGGCTAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAA 1779
       |||:::|||:::   :::   :::   :::      :::|||
 557 leAspTyrAspSerLeuLysArgLysLysAsnSerLeuAsnAlaValPro 573

1780 AGTAGAATAATGCTTAAGGATAAT......................... 1803
        |||::::::     |||:::
 574 ProArgLeuValGluAspAspGlnAspValTyrAspAspValAlaGluGl 590

1804 ..........AAATGGCAGGGTACAAGTAAATCGGGTATTAAAATAGAAG 1843
                 :::    :::|||    |||    ||||||          |
 590 nAspAlaProAsnSerHisGlyGlnSerGlySerGly...........G 603

1844 GATTTACCGAACCTAATAGAACAGCATATCCCATTTATGAA 1884
     |||:::        |||       |||         ||||||:::
 603 lyMetPheProProProProThrAspAspGluIleTyrAsp 616
```

seq_name: SwissProt_40:DAN4_YEAST

seq_documentation_block:

```
ID   DAN4_YEAST     STANDARD;      PRT;  1161 AA.
AC   P47179;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Cell wall protein DAN4 precursor.
GN   DAN4 OR YJR151C OR J2223.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Scarcez T.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REGULATION.
RX   MEDLINE=21113168; PubMed=11160904;
RA   Cohen B.D., Sertil O., Abramova N.E., Davies K.J., Lowry C.V.;
RT   "Induction and repression of DAN1 and the family of anaerobic
RT   mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT   complex array of regulatory sites.";
RL   Nucleic Acids Res. 29:799-808(2001).
CC   -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SRP1 / TIP1 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z49651; CAA89684.1; -.
DR   SGD; S0003912; DAN4.
DR   InterPro; IPR000992; SRP1_TIP1.
DR   Pfam; PF00660; SRP1_TIP1; 1.
DR   PROSITE; PS00724; SRP1_TIP1; 1.
KW   Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1146       CELL WALL PROTEIN DAN4.
FT   PROPEP     1147   1161       REMOVED IN MATURE FORM (POTENTIAL).
FT   LIPID      1146   1146       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   1161 AA;  118358 MW;  7954C15D69F0CA58 CRC64;
```

alignment_scores:

```
           Quality:  153.50            Length:    429
             Ratio:  0.717               Gaps:     12
 Percent Similarity:  49.883   Percent Identity: 22.611
```

alignment_block:
US-09-303-518D-463 x DAN4_YEAST ..

Align seg 1/1 to: DAN4_YEAST from: 1 to: 1161

```
 77 ATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTC 126
    |||:::   :::||||||   |||           :::|||::::::   ||
117 IleTyrThrAlaIleProThrSerThrSerThrThrThrThrLysSerSe 133

127 GAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNA 176
    |:::    |||       ||||||       ::::::       |||      |
133 rThrSerThrThrProThrThrThrIleThrSerThrThrSerThrThrS 150

177 GCGCAACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCC 226
    ||:::|||       ::::::      :::               :::   :::
150 erThrThrProThrThrSerThrThrSerThrThrProThrThrSerThr 166

227 ACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTC 276
    |||                                 ::::::|||   ||
167 Thr..................SerThrThrProThrThrSerThrThrSe 177

277 CGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGC 326
    |:::    |||    |||:::|||:::|||        |||:::|||:::
177 rThrThrProThrThrSerThrThrSerThrThrProThrThrSer.... 192
```

```
 327 CTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCC 376
                         |||      :::::|||    |||:::::::::
 193 .................ThrThrSerThrThrProThrThrSerThrThr 203

 377 TTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTAT 426
        |||:::          :::   |||:::|||    |||     |||:::
 204 SerThrThrProThrThrSerThrThrSerThrThrProThrThrSerTh 220

 427 GACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATAT 476
      |||:::          ::::::           |||
 220 rThrSerThrThrProThrThrSerThrThrProThr............. 232

 477 ATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGA 526
        |||:::||||||        :::                :::    ||||||
 233 ..ThrSerThrThrSerThrThrSerGlnThrSerThrLysSerThrThr 248

 527 CCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCC 576
     |||||||||::::::    ::::::       ||||||:::||||||    ||
 249 ProThrThrSerSerThrSerThr...ThrProThrThrSerThrThrPr 264

 577 GGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCG 626
     |:::                          :::|||    |||:::||||||
 264 oThrThr..................SerThrThrSerThrAlaProThrT 275

 627 ATACAGCCCCGAGCTGGACAGATCGG............GCAATGCCGCCG 664
      ::::::    |||    |||:::                  :::      |||
 275 hrSerThrThrSerThrThrSerThrThrSerThrIleSerThrAlaPro 291

 665 AAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCA 714
             ||||||:::    :::        |||     :::|||::::::    ::
 292 ThrThrSerThrThrSerSerThrPheSerThrSerSerAlaSerAlaSe 308

 715 GGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTC 764
      ::::    :::|||:::    |||                :::    |||:::
 308 rSerValIleSerThrThrAlaThrThrSerThrThrPheAlaSerLeuT 325

 765 AAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGA 814
       |||        :::    ||||||                      |||
 325 hrThrProAlaThrSerThrAla.........SerThrAspHisThrThr 338

 815 TGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCC 864
          :::||||||           :::        ::::::    |||
 339 SerSerValSerThrThrAsnAlaPheThrThrSerAlaThrThrThrTh 355

 865 GCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGG 914
        :::    :::    |||::::::|||:::          :::        ::::
 355 rThrSerAspThrTyrIleSerSerSerSerProSerGlnValThrSerS 372

 915 CATAGAAGCCGTCAGCAATA............TCTTTATGGCAGCCATCC 952
      ::    :::|||:::::::::                |||        :::|||:::
 372 erAlaGluProThrThrValSerGluValThrSerSerValGluProThr 388

 953 CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG..... 997
         |||:::    :::         |||:::|||    |||:::    :::
 389 ArgSerSerGlnValThrSerSerAlaGluProThrThrValSerGluPh 405

 998 ......................GCATCACGGCACATCCTGTCAAGCGGTC 1025
                              |||     ::::::    ||||||::::
 405 eThrSerSerValGluProThrArgSerSerGlnValThrSerSerAlaG 422

1026 GCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACA 1075
      ::        :::    |||:::           ::::::       |||:::    :::
 422 luProThrThrValSerGluPheThrSerSerValGluProThrArgSer 438

1076 ATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGA 1125
                  :::    |||    :::::::::|||    |||:::    ||
 439 .........SerGlnValThrSerSerAlaGluProThrThrValSerGl 452

1126 AATA......TCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACAT 1169
      |:::           ||||||:::            |||        :::    :::::::
 452 uPheThrSerSerValGluProThrArgSerSerGlnValThrSerSerA 469

1170 CACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAG 1219
          ::    |||                :::    |||::::::::::::      :::
 469 laGluProThrThrValSerGluPheThrSerSerValGluProThrArg 485

1220 ACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCG 1269
      ::::::    :::          :::|||             |||||||:::    |||
 486 SerSerGlnValThrSerSerAlaGluProThrThrValSerGluPhe.. 501

1270 AATTTTGAGAAGCACGTGAAATATGATACGAAGCTCG 1306
                  |||:::    :::    |||||||||||
 502 .......ThrSerSerValGluProIleArgSerSer 511
```

seq_name: SwissProt_40:AMYH_YEAST

```
seq_documentation_block:
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
```

```
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN      210   1367       SER/THR-RICH.
FT   CARBOHYD    817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;

alignment_scores:
          Quality:  146.00            Length:    438
            Ratio:  0.737               Gaps:     16
 Percent Similarity:  45.205   Percent Identity: 23.288

alignment_block:
 US-09-303-518D-463 x AMYH_YEAST   ..

 Align seg 1/1  to: AMYH_YEAST  from: 1  to: 1367

      20 TATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGC......CGATGCAT 63
         :::|||   |||      |||||||:::        |||      :::|||
     192 PheProGlyPhe......TyrTrpAsnIleAspCysAspAsnAsnCysGl 206

      64 GCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGA 113
            |||                               :::   ||||
     206 yGlyThr..............................LysSerSerT 212

     114 CCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGG 163
         ||::::::   |||:::        ::::::||||||   |||:::::::::
     212 hrThrThrSerSerThrSerGluSerSerThrThrThrSerSerThrSer 228

     164 GGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGC 213
            |||       :::::::::::   |||
     229 GluSerSerThrThrThrSerSerThrSer.................... 238

     214 CATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATAT 263
                                                 |||   |
     239 ..........................................GluSerS 241

     264 CGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCT 313
         ||:::|||   |||:::          |||:::   :::|||
     241 erThrThrThrSerSer.........ThrSerGluSerSerThrSerSer 254

     314 TCGACAACCATGCCTCACATTCCGATTCTGACG..............AA 348
         ||||||||   |||       |||       |||              ||
     255 SerThrThrAlaProAlaThrProThrThrThrSerCysThrLysGluLy 271

     349 GCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGG 398
         ||||:::   |||   |||   |||:::                  |||
     271 sProThrProProThrThrThrSerCys.................ThrL 282

     399 ATACGAACACCATCCCGCCG............................ 418
            :::       ||||||
     282 ysGluLysProThrProProHisHisAspThrThrProCysThrLysLys 298

     419 .........ACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCC 459
                  |||:::   |||       :::   ::::::   |||:::||
     299 LysThrThrThrSerLysThrCysThrLysLysThrThrThrProValPr 315

     460 AAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAA 509
         |          :::       |||   ::::::          :::|||
     315 oThrProSerSerSerThrThrGluSerSerSerAlaProValProThrP 332

     510 TATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCG 559
           |||:::||||||       :::::::::||||||   ::::::
     332 roSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 348

     560 ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGA 609
         |||   |||:::           :::         ::::::   :::|||::
     349 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGl 365

     610 TTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
         :|||::::::   |||   |||:::   |||   |||:::   :::
     365 uSerSerSerAlaProValThrSer...SerThrThrGluSerSerSerA 381

     660 CGCCGAAGCCTTCAACGGCACTGCAGATATCGTCA...........AAA 697
           |||     |||::::::   :::   ||||||
     381 laProValThrSerSerThrThrGluSerSerSerAlaProValProThr 397

     698 ACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAG 747
            ||||||:::         |||:::   ||||||      :::
     398 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerTh 414

     748 GGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCT 797
          :::   :::::::::          :::   |||   |||:::
     414 rThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerS 431

     798 TTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGC 847
            |||   |||            :::||||||
     431 erAlaProValThr.........SerSerThr..............Thr 439

     848 AACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAAC 897
         :::|||::::::   |||          |||:::       :::|||:::::
     440 GluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAl 456

     898 CCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATA.........TCTT 938
         :|||:::|||               ||||||:::              |||
     456 aProValPro...........ThrProSerSerSerThrThrGluSerS 469

     939 TATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACG 988
               |||      |||:::       |||   |||:::        :::
     469 erSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 485

     989 GCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCG 1038
                :::|||          ||||||:::         :::
     486 ProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSe 502

    1039 ATCGCATTGCCGAAAGGGAAATCCGCCG.................TCAG 1070
            |||      ::::::::::::   |||                  |||:
     502 rSerThrThrGluSerSerSerAlaProValProThrProSerSerSerT 519

    1071 CGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATT 1120
         :::|||                :::|||:::                |||
     519 hrThrGluSerSerSerAlaProAlaProThrProSerSerSerThrThr 535

    1121 CCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATC 1170
            :::   |||:::   :::   ||||||:::|||   :::::::::::
     536 GluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAl 552

    1171 ACCT...CCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGC 1217
         :|||   |||   |||                |||   :::   |||:::
     552 aProValProThrProSerSerSerThrThrGluSerSerSerThrProV 569

    1218 AGACCAACGCCACC 1231
            |||:::::|||
     569 alThrSerSerThr 573

seq_name: SwissProt_40:NKCR_MOUSE

seq_documentation_block:
ID   NKCR_MOUSE     STANDARD;      PRT;  1453 AA.
AC   P30415;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE   related protein) (NK-TR protein).
GN   NKTR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93133824; PubMed=8421688;
```

```
RA   Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA   Ortaldo J.R.;
RT   "A cyclophilin-related protein involved in the function of natural
RT   killer cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN   [2]
RP   REVISIONS TO C-TERMINUS.
RC   STRAIN=BALB/C; TISSUE=Blood;
RA   Anderson S.K.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC       INVOLVED IN THE FUNCTION OF NK CELLS.
CC   -!- SIMILARITY: CONTAINS 1 CYCLOPHILIN-LIKE PPIASE DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L04289; AAA37500.2; ALT_INIT.
DR   HSSP; Q27450; 1A33.
DR   MGD; MGI:97346; Nktr.
DR   InterPro; IPR002130; CSA_PPIase.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT   DOMAIN        1    176       PPIASE, CYCLOPHILIN-TYPE.
FT   DOMAIN      222    241       ARG/LYS-RICH (BASIC).
FT   DOMAIN      422    459       ARG/LYS-RICH (BASIC).
FT   DOMAIN      964   1003       ARG/LYS-RICH (BASIC).
FT   DOMAIN      198    273       ARG/SER-RICH.
FT   DOMAIN      468    565       ARG/SER-RICH.
FT   DOMAIN      658    812       ARG/SER-RICH.
FT   DOMAIN     1303   1453       ARG-SER TANDEM REPEAT-RICH.
SQ   SEQUENCE   1453 AA;  163439 MW;  DF1173FF814B283E CRC64;

alignment_scores:
           Quality:  143.50            Length:    507
             Ratio:  0.595               Gaps:     29
 Percent Similarity:  47.535   Percent Identity:  25.049

alignment_block:
 US-09-303-518D-463 x NKCR_MOUSE   ..

 Align seg 1/1  to: NKCR_MOUSE  from: 1  to: 1453

     304 CATTCGCCCTTCGACAACCAT......GCCTCACATTCCGATTCTGACGA 347
         ||||||      |||:::::      |||||||||  ||||||::
     513 HisSerSerArgAspSerTyrArgSerLysSerHisSerArgSerAspSe 529

     348 AGCCGGTAG.TCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGAC 396
          :  |||  ||||||  |||
     529 rArgGlySerSerArgSerArg..........................A 537

     397 GGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGG 446
            :::          |||  |||  |||  |||:::  :::|||
     537 laValSerLysSerSerSerArgSerLeuAsnArgSerLysSerArgSer 553

     447 CTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAG 496
             |||||||||      |||:::   |||:::
     554 SerSerArgSerGlyProArgArgThrSerIleSer.............. 565

     497 GCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGA 546
                ||||||  |||  |||  :::::::::          :::|
     566 .......ProLysLysProAlaGlnLeuSerGluAsnLysProValLysT 580

     547 CAACGGCTTGCCGACCG...TTTCCACAATGCCGGCGCTATGCTGACGCA 593
         ||      ||||||     ||||||    |||    :::|||:::||
     580 hrGluProLeuArgProSerValProGln.AsnGlyAsnValLeuValGl 596

     594 AGGAGTAGGCGACGGAT....................TCAAACGCGCCA 622
         |   |||  |||::::::                   |||::::::|||
     596 nProVal.AlaAlaGluAsnIleProValIleProLeuSerAspSerPro 612

     623 CCCGATACAGCCCCGAGC...TGGACAGATCGGGCAATGCCGCCGAAGCC 669
         |||        ||||||  |||                |||  |||||
     613 Pro.........ProSerArgTrpLysProGlyGlnLysProTrpLysPr 626

     670 TTCA......ACGGCACTGCAGATATCGTCAAAAACA............. 700
         ||||              :::::::::::||||||
     626 oSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHisLeuLeuP 643

     701 ........TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCG 742
                 |||:::           :::::::::|||    :::
     643 roValGlnSerThrTyrSerLeuThrAsnIleLysAlaThrVal.SerSe 659

     743 TGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 792
          :::::|||   :::|||   |||                    ||||
     659 rSerSerTyrHisLysArgGluLysPro......SerGluSerAspGlyS 674

     793 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATAT 842
         ||||||:::   :::   :::|||    :::              ||||||
     674 erAlaTyrSerLysTyrSerAspArgSerSerGlySerSerGlyArgSer 690

     843 GGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCC 892
         |||:::   :::       |||||||||
     691 GlySerLysSerSerArgSerArgSerSer.................... 700

     893 AAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATG 942
                    |||:::      ||||||||||:::::    |||
     701 ..........SerArgSerTyrThrArgSerArgSerArgSerLeuProT 714

     943 GCAGCCAT..............CCCCATCAAAGGGATTGGAGCTGTCCG 977
           |||:::             |||   :::|||:::   |||   |||
     714 hrSerArgSerLeuSerArgSerProSerSerArgSerHisSer...Pro 729

     978 GGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGC 1027
         :::|||           |||   :::|||:::   |||    :::    :::::
     730 AsnLysTyrSerAspGlySerGlnHisSerArgSerSerSerTyrThrSe 746

    1028 A...........GATGGGCGCGATCGCATT......GCCGAAAGGGAAA 1059
         :            |||||||||      :::      ::::::||| |||
     746 rValSerSerAspAspGlyArgArgAlaMetPheArgSerAsnArg.Lys 762

    1060 TCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTC 1109
         ::::::|||:::::::::             :::   ::::::       ::
     763 LysSerValThrSerHisLysArgHisArgSerAsnSerGlu......Ly 777

    1110 CCCTTACCATTCCCGAAATATCCGTTCA...AACTTGGAGCAGCGTTACG 1156
         :      ||||||:::   :::|||:::   :::   ::::::|||:::
     777 sThrLeuHisSerLysTyrValArgGlyArgGluLysSerSerArgHisA 794

    1157 GCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAAT 1206
            |||   :::   :::    |||::::::          ::::::
     794 rgLysTyrSerGluSerArgSerSerLeuAspTyrThrSerAsp...... 808

    1207 GTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGG 1256
                  :::||||||   |||   ::::::
     809 .........SerAspGlnSerHisValGlnVal................ 816

    1257 TAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCG 1306
              :::   |||:::   ||||||:::   |||   ::::::      |
     817 .TyrSerAlaProGluLysGluLysGlnGlyLysValGluAlaLeuAsnA 833

    1307 ATATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTT 1356
         ||   |||              |||   |||          ::::::||||||
     833 spLysGln.........GlyLysGlyArgGluGluGlyLysPro...... 844

    1357 GATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAAC 1406
                |||||||:::||||||       |||
```

```
 845 ......LysProGluTrpGluCysProArg................... 852

1407 TCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGA 1456
             :::   :::::|||   ::::::      |||   ::::::
 853 .......SerLysLysGluAsnSerGluAspHisSerArgAspAspSerV 867

1457 GTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAAAACAGGG 1506
       |||:::   |||      |||      :::|||::::::::::::
 867 alSerLysGlyLysAsnCysAlaGlySerLysTrpAspSerGluSer... 882

1507 TTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGT 1556
                                          |||::::::     ||
 883 ...............................AsnSerGluGlnAspVa 888

1557 AACAGGAGGGCATAGTCTA......ACCCGTGGTGATGTACGGGTGATAC 1600
      ||||   :::::::::            |||||||:::   :::
 888 lThrLysSerArgLysSerAspProArgArgGlySerGluLysGluGluG 905

1601 AACAAACCTCGGCACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAA 1650
        ::::::|||:::   ::::::       |||   |||:::   ::::::
 905 lyGluAlaSerSerAspSerGluSerGluValGlyGlnSerHisIleLys 921

1651 ATTAAAAAGCCT..................................GA 1664
        |||   |||                                  ||
 922 AlaLysProProAlaLysProProThrSerThrPheLeuProGlySerAs 938

1665 TGGAAGTTGGGAG 1677
      ||||:::|||:::
 938 pGlyAlaTrpLys 942
```

seq_name: SwissProt_40:YAV1_SCHPO

seq_documentation_block:

```
ID   YAV1_SCHPO     STANDARD;      PRT;  1794 AA.
AC   Q10172;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 193.3 kDa protein C27F1.01C in chromosome I.
GN   SPAC27F1.01C OR SPAC25G10.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE OF 1-1748 FROM N.A.
RC   STRAIN=972;
RA   Harris D., McDonald S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1457-1794 FROM N.A.
RC   STRAIN=972;
RA   McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: SOME, TO YEAST PAN1 AND TO MAMMALIAN EPS15.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; Z69368; CAA93290.1; -.
DR   EMBL; Z70691; CAA94638.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR000261; EPS15_repeat.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF00036; efhand; 3.
DR   Pfam; PF02205; WH2; 1.
DR   SMART; SM00027; EH; 2.
DR   SMART; SM00246; WH2; 1.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE   1794 AA;  193279 MW;  1F042BA5E80E9E27 CRC64;
```

alignment_scores:

```
          Quality:  143.50            Length:    420
            Ratio:  0.829               Gaps:     20
Percent Similarity:  41.190   Percent Identity: 22.143
```

alignment_block:

US-09-303-518D-463 x YAV1_SCHPO ..

Align seg 1/1 to: YAV1_SCHPO from: 1 to: 1794

```
  12 CCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATGC 61
      |||:::::        |||   :::|||:::   :::|||   :::||
1326 ProSerThrThrSerThrSerPheAsnThrAlaProIleProGlnGlnAl 1342

  62 ATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTC 111
      |       :::   :::|||:::|||   ||||||   |||
1342 aProLeuGluAsnGlnPheSerLysMetSerLeuGluPro.......... 1355

 112 GACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAG 161

1355 .................................................. 1355

 162 GGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAA 211
                             |||   |||   :::     |||::
1356 .......................ProValArgProAlaValProThrSe 1364

 212 GCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAAT 261
      :|||:::         ||||||:::|||
1364 rProLysProGlnIleProAspSerSer...................... 1373

 262 ATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCC 311
        ::::::|||:::|||      |||                :::::|
1374 ..AsnValHisAlaProPro..Pro..............ProValGlnP 1384

 312 CTTCGACAACCATGCCT...CACATTCCGATTCTGACGAAGCCGGTAGTC 358
      ||    ::::::||||||   |||      :::   ::::::|||   :::
1384 roMetAsnAlaMetProSerHisAsnAlaValAsnAlaArgProSerAla 1400

 359 CCGTTGACGGATTCAGCCTTT........................ACCGC 384
      |||           ::::::|||                     :::::
1401 ProGluArgArgAspSerPheGlySerValSerSerGlySerAsnValSe 1417

 385 ATCCATTGGGACGGATACGAACACCATCCCG.................. 415
      :||||||       :::|||:::|||:::|||
1417 rSerIleGluAspGluThrSerThrMetProLeuLysAlaSerGlnProT 1434

 416 .....CCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCG...... 454
           |||   |||   :::::::|||           :::::::|||
1434 hrAsnProGlyAlaProSerAsnHisAlaProGlnValValProProAla 1450

 455 .......................CTCCCAAAGGCGCGAGGGATATATAC 480
                             |||||||||   |||:::   ||
1451 ProMetHisAlaValAlaProValGlnProLysAlaProGlyMetValTh 1467

 481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGA 530
      |     :::                         :::   |||:::   |||:
1467 rAsnAlaPro....................AlaProSerSerAlaProA 1477

 531 CAACCGCAGCACCGGACAACGGCTTGCCGACCGTT............... 565
      ::      ||||||   :::   ||||||   :::
1477 laProProAlaProValSerGlnLeuProProAlaValProAsnValPro 1493

 566 ......TCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGA 609
            |||   :::|||:::::::::   :::
1494 ValProSerMetIleProSerValAlaGlnGln...........ProPr 1506

 610 TTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
      |||::::::   |||   |||||||||||   |||          :::|
```

```
1506 oSerSerValAlaProAlaThrAlaProSerSerThr.........LeuP 1520

 660 CGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCG 709
     ||||:::   |||:::      ::::::  |||   :::
1520 roProSerGlnSerSerPheAlaHisValProSerProAla......... 1533

 710 CGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAA 759

1533 .................................................. 1533

 760 GGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAA 809
                                              ||||||
1534 ........................................ProPro.... 1535

 810 CAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACT 859

1535 .................................................. 1535

 860 ATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCA........ 901
        |||||:::||||||||   :::   |||:::::|||
1536 AlaProGlnHisProSerAlaAlaAlaLeuSerSerAlaProAlaAspAs 1552

 902 ....ATGCCGCACAAGGCA.........................TAGA 920
         |||||||||:::::                          |
1552 nSerMetProHisArgSerSerProTyrAlaProGlnGluProValGlnL 1569

 921 AGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 970
     |||||:::|||||:::           |||:::  :::  |||   :::
1569 ysProGlnAlaIleAsnAsnIleAlaProAlaThrAsnLeuGlyThrSer 1585

 971 CTGTCCGGGGAAAATACGGCTTGGG......CGGCATCACGGCACATCCT 1014
        |||  :::    |||:::|||      ::::::  |||::: |||
1586 GlnSerPheSerPro.ArgMetGlyProValAsnAsnSerGlySer.Pro 1601

1015 GTCAAGCGGTCGCAGATGGGC......GCGATCGCATTGCCGAAAGGGAA 1058
     :::      :::      |||      ::::::|||:::|||
1602 LeuAlaMetAsnAlaAlaGlyGlnProSerLeuAlaValProAlaValPr 1618

1059 ATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGT 1108
      ||||||   |||::::::|||          |||     :::  |||:
1618 oSerAlaProSerAsnHisPheAsnProPheAlaLysMetGlnProProA 1635

1109 CCCCT 1113
      ::|||
1635 laPro 1636
```

seq_name: SwissProt_40:KP58_DROME

```
seq_documentation_block:
ID   KP58_DROME     STANDARD;      PRT;   952 AA.
AC   Q9VPC0; Q94889; Q9TXB3;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Serine/threonine protein kinase PITSLRE (EC 2.7.1.37) (Cell division
DE   cycle 2-like).
GN   PITSLRE OR CG4268.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND TISSUE SPECIFICITY.
RC   TISSUE=Embryo;
RX   MEDLINE=97084555; PubMed=8930898;
RA   Sauer K., Weigmann K., Sigrist S., Lehner C.F.;
RT   "Novel members of the cdc2-related kinase family in Drosophila:
RT   cdk4/6, cdk5, PFTAIRE, and PITSLRE kinase.";
RL   Mol. Biol. Cell 7:1759-1769(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE OF 571-722 FROM N.A.
RC   TISSUE=Imaginal disks;
RX   MEDLINE=92335284; PubMed=1378625;
RA   Biggs W.H. III, Zipursky S.L.;
RT   "Primary structure, expression, and signal-dependent tyrosine
RT   phosphorylation of a Drosophila homolog of extracellular signal-
RT   regulated kinase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:6295-6299(1992).
CC   -!- FUNCTION: Acts as a negative regulator of the normal cell cycle
CC       progression. May function in regulating proliferation by the
CC       phosphorylation and subsequent plasma membrane targeting of
CC       galactosyltransferase (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Present throughout the early embryo. In
CC       late embryos levels are highest in the CNS.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC       Highest levels in early embryogenesis (0-6 hours), low levels
CC       during later embryogenesis, moderate levels in pupae and adults.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       CDC2/CDKX SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X99513; CAA67863.1; -.
DR   EMBL; AE003592; AAF51635.1; -.
DR   FlyBase; FBgn0016696; Pitslre.
```

```
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_kin_actsite.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;
KW   Nuclear protein.
FT   DOMAIN       51     56       NUCLEAR LOCALIZATION SIGNAL
FT                                (POTENTIAL).
FT   DOMAIN       92    146       HIS-RICH.
FT   DOMAIN      386    504       SER-RICH.
FT   DOMAIN      558    851       PROTEIN KINASE.
FT   NP_BIND     564    572       ATP (BY SIMILARITY).
FT   BINDING     587    587       ATP (BY SIMILARITY).
FT   ACT_SITE    685    685       BY SIMILARITY.
FT   CONFLICT     69     69       D -> E (IN REF. 1).
FT   CONFLICT    223    223       V -> A (IN REF. 1).
FT   CONFLICT    283    283       I -> S (IN REF. 1).
FT   CONFLICT    492    492       V -> A (IN REF. 1).
FT   CONFLICT    584    584       V -> E (IN REF. 3).
FT   CONFLICT    603    603       L -> R (IN REF. 3).
FT   CONFLICT    612    612       G -> GG (IN REF. 3).
FT   CONFLICT    670    670       MISSING (IN REF. 3).
FT   CONFLICT    680    680       MISSING (IN REF. 3).
FT   CONFLICT    717    717       MISSING (IN REF. 3).
SQ   SEQUENCE   952 AA;  108837 MW;  9CBDE8D459D0713D CRC64;

alignment_scores:
          Quality:  141.00              Length:    473
            Ratio:  0.662                 Gaps:     25
 Percent Similarity:  45.032    Percent Identity: 21.353

alignment_block:
 US-09-303-518D-463 x KP58_DROME   ..

 Align seg 1/1  to: KP58_DROME  from: 1  to: 952

  393 GGACGGATACGAACACCA.................TCCCGCCGACGGCT 424
      |||:::::|||:::|||                  ||||||
   10 GlyGlnLeuArgSerProAsnAspValHisTyrHisSerArg........ 23

  425 ATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGA. 473
                               ::::::::::::::::||||||
   24 ........................GlyGluGluAspGluHisGluGlyA 32

  474 .............TATATACAGCTACGACATAAAAGGCGTTGCCCAAAA 509
                   |||||||||
   32 spAlaAspAlaLeuTyrIleGln.......................... 39

  510 TATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCG 559
         |||||||||   :::|||:::
   40 ...ProProGlnAlaSerArgGluSerGlySerGly............. 50

  560 ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGA 609
             |||:::   ::::::          :::|||:::|||||||||
   51 .......ProArgArgGluLysLysLysHisSerArgGluArgArgArgH 65

  610 TTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
        :::    |||                    :::::::|||   :::
   65 isLysGluArg....................AspAspValGlyGlyAla 74

  660 CGCCGAAGCCTTCAACGGCACTGCAGATAT...........CGTCAAAA 697
       :::|||:::|||      ||||||            |||:::::
   75 AlaLeuAlaLeuGluArgAspHisArgTyrAspTyrArgSerArgGluGl 91

  698 ACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAG 747
       :|||:::::::::   :::|||:::   |||      ::::::   |||:
   91 uHisTyrHisHisHisGlnArgGlu...ArgSerSerAsnAlaAlaAlaA 107

  748 GGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCT 797
      :::|||          ||||||            :::||||||   |||
  107 laTyr.........AlaLysHis............HisLeuGlyHisAla 116

  798 TTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGC 847
       :::|||       |||          :::|||:::
  117 TyrHisTyrProGlnProProGlnGlnGlnGln................ 127

  848 AACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAAC 897
          |||   |||                             |||   |
  128 ....GlnProLeu.............................ProProAlaP 134

  898 CCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGC 947
      ||:::       ::::::|||       ::::::|||:::|||   |||:::
  134 roSerTyrAlaAlaHisHisTyrHisHisHisGlnHisLeuSerGlyAla 150

  948 CATCCCCATCAAAGGGATTGG...AGCTGTCCGGGGAAAATACGGCTTGG 994
      :::          |||::::::   |||   |||:::      :::   ||
  151 ArgAlaAlaProArgGluTyrHisSerTyrProSerGlyTyrHisSerGl 167

  995 GCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCA 1044
      |   :::||||||                                |||
  167 ySerArgHisGly..............................AspTyrP 174

 1045 TTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATA 1094
         |||:::       ||||||  |||:::::::::|||::::::
  174 roMetGluGluProThrArgArg.SerSerLysTyrAlaGluSerLysAs 190

 1095 CGCCAAA.......................................... 1101
       |||:::
  190 pAlaGluSerLeuGluGlnAspLeuArgSerArgLeuLeuLysLysArgH 207

 1102 .....TACCCGTCCCCTTACCATTCC...CGAAATATCCGTTCAAACTTG 1143
           |||   :::   |||::::::   :::|||   :::   ::::::
  207 isAsnTyrValLysAspTyrGluThrGluGluAsnTyrGluHisArgVal 223

 1144 GAG.........CAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGT 1184
      |||          :::|||   |||   ::::::      :::    |||||
  224 GluArgSerAspArgArgGluGlyGlyArgLysGluArgGluArgThrVa 240

 1185 GCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGA 1234
      |     ::::::   :::|||                     ||||||   :
  240 lArgSerThrHisLysGlnAsn..................ArgHisAspA 251

 1235 AGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCAC 1284
      :::::             :::|||:::            |||:::   |||   |||
  251 rgValIleGluLeuLeuAspSer.........ProGluGlnGluHisHis 264

 1285 GTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGGTAT 1334
         :::::::::   |||
  265 HisGlnHisGlnHisLys................................ 270

 1335 ACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGG.......... 1374
                               :::   :::   :::|||
  271 ......................SerHisArgSerLysTrpArgGluGluV 280

 1375 .................GAGGTTGATAGGAAGCTTAATAAATTGACAACT 1407
                      :::|||   :::   |||:::   |||:::
  280 alGluValIleArgArgLysValProGluAspLeuGluLeuLeuAlaArg 296

 1408 CGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAG 1457
      ||||||::::::             :::|||:::|||:::|||:::|||:::
  297 ArgGluLysLeuLeuAlaAlaGluArgGluSerArgGlnArgLysGlnTh 313

 1458 TAGTCAGTTTAAAGCCCATGCGCAACGAGAATGG...............G 1492
      ::::::   :::   :::|||:::||||||                   |
  313 rAlaArgGluGluLeuGluAlaArgArgGluLeuLeuArgGluArgAsnG 330

 1493 AAAATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAAT 1542
      ||::::::   :::|||
  330 luHisSerAspAlaLeu................................ 335

 1543 AAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGTGATGTACG 1592
         :::   :::|||   :::   |||:::|||   |||
```

```
  336 ......SerProThrThrValAlaAlaSerValThrAlaGlyLeuAsnIl 350

 1593 GGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTTTATCAAGCGA 1642
          :::::::::   |||   |||||||:::          |||:::
  350 eHisValLysArgLysSerLysProAspAsn.........TyrGluLysG 364

 1643 CAGTGGAAATTAAAAAG 1659
        :::::::::||||||
  364 luIleLysLeuLysLys 369
```

seq_name: SwissProt_40:YGY3_HALSQ

```
seq_documentation_block:
ID   YGY3_HALSQ     STANDARD;      PRT;   437 AA.
AC   P21561;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 50.6 kDa protein in the 5'region of GYRA and GYRB (ORF
DE   3).
OS   Haloferax sp. (strain Aa 2.2).
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX   NCBI_TaxID=2254;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91100352; PubMed=1846146;
RA   Holmes M.L., Dyall-Smith M.L.;
RT   "Mutations in DNA gyrase result in novobiocin resistance in
RT   halophilic archaebacteria.";
RL   J. Bacteriol. 173:642-648(1991).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M38373; -; NOT_ANNOTATED_CDS.
DR   PIR; C39135; C39135.
KW   Hypothetical protein.
SQ   SEQUENCE   437 AA;  50626 MW;  B5B99A2AF3892BEF CRC64;

alignment_scores:
          Quality:  139.00             Length:    439
            Ratio:  0.785                Gaps:     25
 Percent Similarity:  40.319   Percent Identity: 25.285

alignment_block:
 US-09-303-518D-463 x YGY3_HALSQ   ..

 Align seg 1/1  to: YGY3_HALSQ  from: 1  to: 437

  102 GCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTAT 151
       |||   ::::||||||   :::   |||   :::||||:::      |||:::
   17 AlaArgAlaArgProArgThrArgArgArgHisArgAsnAspHisProLe 33

  152 TCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATAT.......... 191
       :                               |||   |||   |||
   33 u....................LeuAlaGlyArgArgTyrLeuArgAspA 43

  192 ........CGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGAT 233
               |||:::                ||||||   :::
   43 spArgValArgLeuGlnAspAlaArgLysProProAla............ 55

  234 GATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTT 283
                                      |||   :::   |||   ::
   56 .............................ArgAlaArgValProGlnLe 62

  284 CCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACAT 333
       :|||:::|||              ||||||||||||:::

   62 uArgGlyArg.........AspPheAlaLeuArgArg............ 71

  334 TCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCG 383
                                :::        |||::::::   ::::|||
   72 ......................AlaAspArgArgValGluHisValPro 80

  384 ............CATCCATTGGGACGGATACGA................. 404
                  ||||||      |||:::|||
   81 LeuArgGlyArgHisPro......ArgValArgArgValProGlnArgAs 95

  405 ..........ACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGC 444
                 :::|||   ||||||:::|||   |||         ||||
   95 pGlnAspGlyAlaProArgArgArgHisLeuLeuArg.........ArgA 109

  445 GGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACAT... 491
       ||::::::     :::   |||:::                   ||||||
  109 rgValGlyGlyHisArgGlyArgAsn..............ArgHisAla 120

  492 ...AAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCA 538
            ||||||:::|||            :::        :::|||   ||
  121 GlyAspArgArgAlaProGlyValAspSerArgLeuArgGlnGlnHisGl 137

  539 GCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTG 588
       ||||                          |||:::   |||:::    :::|
  137 nHis.......................ProArgGlyArgHisAlaSerA 146

  589 AC.......................GCAAGGAGTAGGCGACGGATTCAA 614
       ||                          |||:::|||   |||   |||
  146 spArgValGlnAspGlyAlaHisProArgArgGlnArgLeuArgGluGln 162

  615 ACGCGCCACCCGATACAGCCCCGAGCTGGACAG.........ATCGGGCA 655
          ||||||      :::||||||     |||            |||::
  163 ProArgHisAlaGlyArgProArgArgArgGlnProProArgArgGlyAr 179

  656 ATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATC 705
       :   |||   :::   :::||||||   |||              :::
  179 gSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArgProAlaV 196

  706 GGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAG 755
         ||||||      :::   :::   |||:::   |||:::
  196 alArgGlyProAspGluAspGlnAlaArgGluPheArgGlyProArgHis 212

  756 CGAAGGCTCAAACATTGCTGTCATGCACGG.........CTTGGGTCTGC 796
       ||||||   :::|||        ||||||            |||:::
  213 ArgArgGluArgHisProProThrAlaArgAspValLeuArgGlyGluPr 229

  797 TTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCG 846
          |||          ||||||   |||
  229 oGlyHis......GlyAspGlyHisHis...................... 236

  847 CAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAA 896
                                 :::   |||   |||   |||:::
  237 .......................LeuGluGlyArgArgGlyArgProArg 245

  897 CCCCAATGCCGCACAAGGCATAGAAGC..................CGTCA 928
       ||||||   |||            |||:::                |||  :
  246 ProGlnGlyArgGluAlaGlyArgGlyAlaHisProProGlnValArg.A 262

  929 GCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATT............ 966
       :::::|||::::::||||||       :::|||:::
  262 laArgIleTyrLeuAlaAlaGlyGluAlaArgGlyLeuProGluProArg 278

  967 ...........................GGAGCTGTCCGGGGAAAATA 986
                                   |||    :::||||||:::
  279 ProLeuGlyValArgThrValHisArgGlyGlyArgLeuArgGlyArgVa 295

  987 CGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCG 1036
        |||    :::|||                    |||   |||:::    :
  295 lGlyGlnAlaGly..................ProArgProGlnValProG 306

 1037 CGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT 1086
       ::             |||:::|||:::                          |||
  306 lyAspPheAlaProGlnGlyGlu.......................Asp 314
```

```
 1087 GCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTC 1136
        :::        :::   |||          |||||||||:::    |||::
  315 SerGluArgArgGluThrProProArgProHisSerArgLysArgArgAs 331

 1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGC 1186
       ::::  ||||||                   ||||||        |||
  331 pThr..GlyAla..................HisHisArgHisTrpArgAr 341

 1187 CGCCGTCAAACGGCAAAAATGTCAAACTGGCAGA......CCAACGCCAC 1230
               :::|||   :::   :::   |||            :::   |
  341 gArgArgArgArgValArgHisArgGluGlyAlaLeuProAlaAlaHisP 358

 1231 CCGAAGACAGGCGT 1244
       ||:::|||||||||
  358 roAspAspArgArg 362

seq_name: SwissProt_40:DAB_DROME

seq_documentation_block:
ID   DAB_DROME      STANDARD;      PRT;  2411 AA.
AC   P98081;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Disabled protein.
GN   DAB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93194063; PubMed=7680635;
RA   Gertler F.B., Hill K.K., Clark M.J., Hoffmann F.M.;
RT   "Dosage-sensitive modifiers of Drosophila abl tyrosine kinase
RT   function: prospero, a regulator of axonal outgrowth, and disabled, a
RT   novel tyrosine kinase substrate.";
RL   Genes Dev. 7:441-453(1993).
CC   -!- FUNCTION: TOGETHER WITH ABL INVOLVED IN EMBRYONIC NEURAL
CC       DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: CNS AXONS AND BODY WALL MUSCLES.
CC   -!- DEVELOPMENTAL STAGE: EMBRYONIC AXONOGENESIS.
CC   -!- PTM: PROBABLY PHOSPHORYLATED BY THE ABL TYROSINE KINASE.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L08845; AAB08527.1; ALT_SEQ.
DR   PIR; A46299; A46299.
DR   FlyBase; FBgn0000414; Dab.
DR   InterPro; IPR000050; PID_domain.
DR   Pfam; PF00640; PID; 1.
DR   SMART; SM00462; PTB; 1.
DR   PROSITE; PS01179; PID; 1.
KW   Alternative splicing; Phosphorylation; Repeat.
FT   DOMAIN       46    196       PID.
FT   DOMAIN     1689   1801       REPEAT-RICH REGION.
FT   REPEAT     1689   1700       ALTERNATE ARG AND ACIDIC RESIDUE.
FT   REPEAT     1740   1750       ALTERNATE ARG AND ACIDIC RESIDUE.
FT   REPEAT     1791   1801       ALTERNATE ARG AND ACIDIC RESIDUE.
FT   MOD_RES     111    111       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES     482    482       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1662   1662       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1667   1667       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1701   1701       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1704   1704       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1713   1713       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1739   1739       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1826   1826       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   MOD_RES    1961   1961       PHOSPHORYLATION (BY ABL) (PROBABLE).
FT   VARSPLIC    462    673       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   2411 AA;  264046 MW;  579AB9C0243D5FD6 CRC64;

alignment_scores:
           Quality:  138.50          Length:    558
             Ratio:  0.523             Gaps:     38
 Percent Similarity:  47.491  Percent Identity: 23.835

alignment_block:
 US-09-303-518D-463 x DAB_DROME  ..

 Align seg 1/1  to: DAB_DROME  from: 1  to: 2411

  266 GCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGC...CC 312
       ||||||   ||||||   |||      |||:::   :::::::::   ||
 1405 AlaThrSerSerAlaAlaProSerArgGlyAlaProGlyLeuHisThrPr 1421

  313 TTCG...............ACAACCATGCCTCACATTCCGATTCTGACGA 347
       ||||               :::|||   |||   ||||||   |||   |
 1421 oSerGlnPheAsnAspValSerThrSerPro...IlePro...LeuGlnL 1436

  348 AGCCG...............GTAGTCCCGTTGACGGATTCAGCCTTTACC 382
        |||||             ||||||          ||||||   :::
 1436 ysProGlyMetGlyProSerProValProSerGlnLeuSerAlaValSer 1452

  383 GCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 432
             |||   |||   |||:::                |||   ||
 1453 GlnLeuIleAspThrAlaThrLys...............GlnMetMetGl 1464

  433 CCACAGGGCG..................GCGGCTATCCCGCTCCCAAAGG 464
       |    :::                  |||::::::    ||||||
 1464 yAspLysAspArgGluLysGlnSerTrpAlaThrPheAspSerProLysA 1481

  465 CGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCC 514
       ||:::|||            |||                |||     |||
 1481 laLysGlyLysAlaArgLeuThrLeuProProProProProProAlaSer 1497

  515 GCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCG..... 559
         ::::::   |||   |||            :::|||   |||
 1498 AsnThrSerGlnProAspThr.........GluAsnArgLeuAlaValAr 1511

  560 .............ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGG 596
                   |||::::::         ::::::      |||:::
 1511 gIleProGlyMetThrAlaGlyGlnSerAsnSerValValGlyArgArgA 1528

  597 AGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACA 646
       ::   ||||||   |||::::::        |||   :::|||   |||
 1528 rgSerAlaThrThrSerSerSerSerArgAspLeuSerPro...TrpAsp 1543

  647 GATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAA 696
       |||::::::         |||:::            |||:::::::::
 1544 AspGluThr......ProGluTyrLeuLysArgArgGlnLeuAlaAla.. 1557

  697 AACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCG.... 742
        ::::::   |||:::   ::::::            ||||||   |||
 1558 .AlaGlnMetAlaHisProHisGlnProProMetGlnAlaProProGlnH 1574

  743 .......................TGCAGGGTATAAGCGAAGGCTCAAACA 769
                              |||  :::   ||||||::::::
 1574 isThrAspArgHisGlyTyrTyrMetArg.HisAlaArgArgMetAsnSe 1590

  770 TTGCTGTCAT...........................GCACGGCTTGGGT 792
       |||     :::                           ||||||     :
 1590 rCysAspGluAspTyrAspTyrAspGlyGluPheValAlaArgArgAspG 1607
```

```
 793 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATAT 842
     ::        |||::::::|||               ||||||    ::::::
1607 lnProGlnHisGlnGlnGln...........GlnArgLysPheLysHis 1619

 843 GGCGCAACTCAAAGACTATGCCGCAGC....................... 869
     |||             |||   ||||||
1620 Gly...........LeuSerArgSerArgAspAsnPheGluLeuGluSe 1632

 870 ............AGCCATCCG......CGATTGGGCAGTCCAAAA.... 896
                      ||||||     :::      |||||||:::
1632 rProSerTrpTyrHisHisProAlaHisHisThrTrpSerProGlnGluI 1649

 897 ..CCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATAT......... 935
          |||:::|||:::|||      :::|||      |||
1649 leGluGlnAlaArgValArgSerPheAspArgThrAlaTyrGluArgSer 1665

 936 CTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAAT 985
        ||||||    |||       |||          |||:::::
1666 SerTyrGlyProProProIleTyrAspLysArg......GlyGlnLe 1680

 986 ACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGC 1035
     :|||   |||::::::::|||                    |||   |
1680 uArg...GlyLysTyrArgGly....................AspHisA 1689

1036 GCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGA 1085
     ||||||||      |||||||||      |||:::|||::: :::   ||
1689 rgAspArg......GluArgGluArgAspArgAspArgGlu.TyrArgAs 1703

1086 TGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTT 1135
     |   |||       :::|||         |||       :::|||:::
1703 pTyrAlaArgProSerTyrAspPheAspTyr......GluAsnVal.... 1716

1136 CAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTG 1185
                |||:::|||  |||::::::         :::
1717 .....TyrGluGluArgGlyGlyArgSerProLeu.........AlaTyr 1728

1186 CCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGAC.............. 1221
        |||::::::|||    :::      |||   |||
1729 LysProGlyArgGlyGlyGlyAspTyrLeuTyrAspArgGluArgAspAr 1745

1222 .......CAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGT 1264
              :::|||   |||||||:::        ||||||   :::::::
1745 gAspArgGluArgAspProLysSer.........PheAspArgGluSerL 1759

1265 TTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAA 1314
     ::   ::::::|||:::   ::::::                     :::
1759 euGluSerTyrGluSerAlaThrArg....................Arg 1768

1315 GAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAA 1364
     :::   |||   |||:::         :::       :::|||:::::
1769 ArgArgSerPheGlySerGlyAsnAspValTyrGlySerLeuAspSerAr 1785

1365 ACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGC 1414
     :       ::::::   ||||||:::   ::::::        |||||||
1785 gAspAspTyrArgGlyAspArgGluArgAspArgGluArgAspArgGluG 1802

1415 AGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGT... 1461
     ||::::::      :::::::::   |||:::   :::   :::|||
1802 lnMetLys......ThrArgSerLeuArgLysProThrThrThrSerGly 1816

1462 ........................CAGTTTAAAGCCCATGCGCAACGAGA 1487
                             :::::::::    ::::::::::
1817 LysLeuArgIleSerGlyAspIleAspTyrGluGlnAspSerGluGlnAs 1833

1488 ATGGGAAAATAAAACAGGG.........TTAGATTTTAATCATTTTATAG 1528
      :::::::::::::::|||         |||::::::   :::   :::|
1833 pPheGlnGlnArgSerGlyValArgSerLeuGlnLeuProAsnGlnLeuG 1850

1529 GTGGTGAT...ATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTA 1575
      |||||||    :::       :::::::::|||:::|||   :::   :::
1850 lyGlyAspValValLeuProSerAsnAlaValAlaGlyProGlnArgPhe 1866

1576 ACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGATAAACA 1625
      ::::::                  ::::::   |||:::|||   :::::
1867 AlaLys.................SerSerGlySerValProGlyArgAr 1877

1626 TGGG.....................GTTTATCAAGCGACAGTGGAAATTA 1654
      :|||                    :::   ::::::         :::
1877 gGlyThrProCysLeuValArgAsnLeuGlySerValGlnProAlaLeuL 1894

1655 AAAAGCCTGATGGAAGTTGG 1674
        ||||||   |||:::|||
1894 euLysProLysGlyAspTrp 1900
```

seq_name: SwissProt_40:PODX_RABIT

```
seq_documentation_block:
ID   PODX_RABIT     STANDARD;      PRT;   551 AA.
AC   Q28645;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Podocalyxin-like protein 1 precursor.
GN   PODXL OR PCLP1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE;
RX   MEDLINE=96094343; PubMed=7493982;
RA   Kershaw D.B., Thomas P.E., Wharram B.L., Goyal M., Wiggins J.E.,
RA   Whiteside C.I., Wiggins R.C.;
RT   "Molecular cloning, expression, and characterization of podocalyxin-
RT   like protein 1 from rabbit as a transmembrane protein of glomerular
RT   podocytes and vascular endothelium.";
RL   J. Biol. Chem. 270:29439-29446(1995).
CC   -!- FUNCTION: Functions as an antiadhesin that maintains an open
CC       filtration pathway between neighboring foot processes in the
CC       podocyte by charge repulsion.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Glomerular epithelium cell (podocyte).
CC   -!- PTM: Sialoglycoprotein.
CC   -!- SIMILARITY: BELONGS TO THE PODOCALYXIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U35239; AAC48489.1; -.
KW   Glycoprotein; Signal; Transmembrane.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    551       PODOCALYXIN-LIKE PROTEIN 1.
FT   DOMAIN       22    452       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    453    473       POTENTIAL.
FT   DOMAIN      474    551       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   551 AA;  57040 MW;  E9B8AE168CDFB8C5 CRC64;

alignment_scores:
          Quality:  138.00             Length:    384
            Ratio:  0.754                Gaps:     23
 Percent Similarity:  47.656    Percent Identity: 26.302

alignment_block:
 US-09-303-518D-463 x PODX_RABIT   ..
```

```
Align seg 1/1  to: PODX_RABIT  from: 1  to: 551

    122 ATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTT 171
        :::|||  |||:::   :::            |||       |||
     15 LeuSerProProSerLeuSerGlnGluLysSerProGlnProGlyProTh 31

    172 GCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTT 221
         |||  ||||||:::  |||          |||     |||   :::
     31 rProMetAlaThrSerThrSer.........ThrArgProAlaProAlaS 45

    222 GGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACA 271
         |||                   |||  ::::::::::::|||:::
     45 erAlaPro..............AlaProLysSerSerValAlaAlaSer 56

    272 TTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAAC 321
        :::   |||                          :::   :::|||||
     57 ValProAla.........................GluGlnAsnThrTh 64

    322 CATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATT 371
        |              |||:::   ||||||     :::|||  |||:::|
     64 r............ProMetThrThrLys......AlaProAlaThrGlnS 75

    372 CAGCCTTTACCGCATCCATTGGGACGGATACG...AACACCATCCCGCCG 418
         ||       :::||||||  |||:::::::::   |||:::   |||
     75 erPro...SerAlaSerProGlySerSerValGluAsnSerAlaProAla 90

    419 ACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCG 468
           :::   |||      :::   :::                ||||||:::
     91 GlnGlySerThrThrThrGlnGlnSerLeuSerValThrThrLysAlaGl 107

    469 A..............GGGATATATACAGCT...ACGACATAAAAGGCGT 500
        :                 |||:::    ||||||   :::|||   :::|||
    107 uAlaLysAspAlaGlyGlyValProThrAlaHisValThrGlySerAlaA 124

    501 TGCCC..............AAAATATCCGCCTCAACCTGACCGACAACC 535
          |||                  :::::::::|||:::         ::::::
    124 rgProValThrSerGlySerGlnValAlaAlaGlnAspProAlaAlaSer 140

    536 GCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCG........ 577
           ||||||:::|||           :::::::::|||   |||
    141 LysAlaProSerAsn......HisSerIleThrThrLysProLeuAlaTh 155

    578 ....GCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCAC 623
          |||                 :::   :::||||||   ::::::|||
    155 rGluAlaThrSerGlnAlaProArgGlnThrThrAspValGlyThrProG 172

    624 CCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCG............ 661
           |||||||||       |||:::   :::   |||
    172 lyProThrAlaProProValThrAsnSerThrSerProAspLeuLeuGly 188

    662 ........CCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATC 702
                |||||||||||   :::   |||:::|||      |||:::
    189 HisAlaThrProLysProSerGluGlyProGlnLeuSerPheProThrAl 205

    703 ATC.GGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTA 751
        :   |||:::   |||   ::::::|||:::|||
    205 aAlaGlySerLeuGlyProValThrGlySerGlyThr............. 217

    752 TAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCC 801
                                        |||   |||   ||||||
    218 .............................GlySerGlyThrLeuSer 223

    802 ACC...GAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCA 848
        |||   ::::::|||   |||   ::::::   :::|||:::   |||:::
    224 ThrProGlnGlyLysProAlaThrLeuThrProValAlaSerSerAlaGl 240

    849 ACTCAAAGACTATG......CCGCAGCAGCCATCCGCGATTGGGCAGTCC 892
        : ::::::   |||        |||:::   ||||||      :::   |||
    240 u.ThrGlnGlyMetProSerProMetProProSerProAlaSerProSer 256

    893 AAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATG 942
        ::::::|||:::|||
    257 SerSerProPheProSer.............................. 262

    943 GCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTT 992
         :::|||||||||||   :::|||:::   ||||||   ::::::|||
    263 .SerProSerProSerProAlaLeuGlnProSerGlyProSerAlaAlaG 279

    993 GGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCG 1042
          :::   :::           ::::::||||||
    279 lyThrGluAsp...........ThrThrGlyArg............... 286

   1043 CATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGG.. 1090
                   |||          |||:::|||   |||
    287 ...........GlyProThrSerSerSerThrGluLeuAlaSerThrAl 299

   1091 ....CATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTC 1136
            |||   |||:::|||         |||
    299 aLeuHisGlyProSerThrLeuSerProThr.................. 309

   1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCA......CCTCCTCAA 1180
                  |||:::|||      :::::::|||      ||||||:::
    310 ........SerAlaValArgAspGlnArgValSerCysGlyProProGlu 323

seq_name: SwissProt_40:MAP4_MOUSE

seq_documentation_block:
ID   MAP4_MOUSE     STANDARD;      PRT;  1125 AA.
AC   P27546;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 4 (MAP 4).
GN   MAP4 OR MTAP4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92042100; PubMed=1718985;
RA   West R.R., Tenbarge K.M., Olmsted J.B.;
RT   "A model for microtubule-associated protein 4 structure. Domains
RT   defined by comparisons of human, mouse, and bovine sequences.";
RL   J. Biol. Chem. 266:21886-21896(1991).
CC   -!- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES
CC       MICROTUBULE ASSEMBLY.
CC   -!- TISSUE SPECIFICITY: TESTIS, STRIATED AND CARDIAC MUSCLE.
CC   -!- PTM: PHOSPHORYLATION OF THE PRO-RICH REGION IN THE C-TERMINUS
CC       NEGATIVELY REGULATES MAP-4 ACTIVITY TO PROMOTE MICROTUBULE
CC       ASSEMBLY.
CC   -!- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M72414; AAA16372.1; -.
DR   PIR; B41206; B41206.
DR   MGD; MGI:97178; Mtap4.
DR   InterPro; IPR001084; Tubulin-bind.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PROSITE; PS00229; TAU_MAP; 3.
KW   Microtubules; Repeat; Phosphorylation.
FT   REPEAT      896    926       TAU/MAP MOTIF.
FT   REPEAT      965    995       TAU/MAP MOTIF.
FT   REPEAT      996   1026       TAU/MAP MOTIF.
SQ   SEQUENCE   1125 AA;  117675 MW;  73047432A329AA1D CRC64;

alignment_scores:
          Quality:  138.00            Length:    632
```

```
            Ratio:  0.502              Gaps:      22
 Percent Similarity:  43.513   Percent Identity: 20.411

alignment_block:
 US-09-303-518D-463 x MAP4_MOUSE   ..

 Align seg 1/1  to: MAP4_MOUSE  from: 1  to: 1125

      90 TCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGA 139
          |||:::   ||||||:::::               ::::::|||:::
     543 SerMetAlaProAlaSerAspLeuAlaLeuProLeuGluThrLysValAl 559

     140 AATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCAT 189
                 ||||||:::::::::|||   :::   :::   :::::::|||
     559 aThrValProIleLysAspLysGlyThrValGlnThrGluGluLysProA 576

     190 ATCGGATTGGG..............AAACATACAAAGCCATCAGTTGGG 224
           :::   :::              :::|||        |||:::
     576 rgGluAspSerGlnLeuAlaSerMetGlnHisLysGlyGlnSerThrVal 592

     225 CCACCTGATGATTCAACAGG.CGGCCGTTGAAGGAAATATCGGCTACATT 273
         ||||||      ::::::   :::|||:::|||      :::     ::
     593 ProProCysThrAlaSerProGluProValLys...AlaAlaGluGlnMe 608

     274 GTCCGCTTTTCCGATCACGGGCACAAATTCCATT................ 307
         :|||::::::||||||   :::   :::   :::
     608 tSerThrLeuProIleAspAlaProSerProLeuGluAsnLeuGluGlnL 625

     308 ..CGCCCTTCGACAACCATGCCTCACATTCCGATTCTG............ 343
           :::        ::::::|||       |||
     625 ysGluThrProGlySerGlnProSerGluProCysSerGlyValSerArg 641

     344 .........ACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGC 384
                  :::|||   :::|||   :::|||   :::   :::|||
     642 GlnGluGluAlaLysAlaAlaValGlyValThrGlyAsnAspIleThr.. 657

     385 ATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCC 434
                  |||      |||     ||||||
     658 .........ThrProProAsnLysGluProPro................ 665

     435 ACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCT 484
                          |||   |||:::   ::::::         |||
     666 ................ProSerProGluLysLysAlaLysProLeuAla 676

     485 ACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAAC 534
         ||||||      |||      :::   |||   :::           ||
     677 ThrThrGlnProAlaLysThrSerThrSerLysAlaLysThrGlnProTh 693

     535 CGCAGCACCGGACAACGGCTTGCCGACCGTTTCC................ 568
         |:::   |||   :::      ||||||:::|||
     693 rSerLeuProLysGlnProAlaProThrThrSerGlyGlyLeuAsnLysL 710

     569 ......................ACAATGCCGGCGCTATGCTGACGCAAG 595
                               ::::::||||||         ::::::
     710 ysProMetSerLeuAlaSerGlySerValProAlaAlaProHisLysArg 726

     596 GAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGAC 645
                |||:::   ::::::|||   |||:::|||   |||:::
     727 ProAlaAlaAlaThrAlaThrAlaArgProSerThrLeuProAla..... 741

     646 AGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCG.... 691
             |||   :::   |||||||||   |||      :::::::::
     742 ....ArgAspValLysProLysProIleThrGluAlaLysValAlaGluL 757

     692 .............TCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCG 727
                      ||||||   |||||||||      |||   :::
     757 ysArgThrSerProSerLysProSerSerAlaProAlaLeuLysProGly 773

     728 GCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTC 777
           ::::::   |||         |||   :::   :::   |||
     774 ProLysThrThrProThrValSerLysAlaThrSerProSerThrLeuVa 790

     778 ATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAA 827
            |||:::                    |||::::::|||      |||:::
     790 lSerThrGly...................ProSerSerArgSerProAlaThr. 801

     828 CGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCC 877
                                          |||:::|||::::::|||:::
     802 ...........................ThrLeuProLysArgProThr 808

     878 GCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTC 927
         :::|||      :::      |||         |||
     809 SerIleLysThrGluGlyLysProAlaAspValLys.............. 820

     928 AGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCG 977

     820 .................................................. 820

     978 GGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGC 1027
           :::   ||||||   :::||||||        ||||||   ::::::
     821 ..ArgMetThrAlaLysSerAlaSerAlaAsp...LeuSerArgSerLys 835

    1028 AGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAAT 1077
               :::   |||      ::::::::::::   |||:::::::::
     836 ThrThrSerAlaSerSerValLysArgAsnThrThrProThrGlyAlaAl 852

    1078 TTTGCCGATGCGGCATACG...CCAAATACCCGTCCCCTTACCATTCCCG 1124
            |||         |||      :::::::::|||::::::   |||:
     852 aProProAlaGlyMetThrSerThrArgValLysProMetSerAlaProS 869

    1125 AA..............ATATCCGTTCAAACTTGGAGCAGCGTTACGGCA 1159
         ::              :::||||||:::         :::   |||
     869 erArgSerSerGlyAlaLeuSerValAspLysLysProThrSerThrLys 885

    1160 AAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTC 1209
            ::::::|||   |||:::         |||      |||
     886 ProSerSerSerAlaProArgValSerArgLeuAlaThr........... 898

    1210 AAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAA 1259
                                                  ||||||:
     899 .........................................ThrValS 901

    1260 AGGGTTTCCGAATTTTGAGAAG...CACGTGAAATATGATACGAAGCTCG 1306
         ::::   |||::::::::::::   :::   |||      :::      :
     901 erAla..ProAspLeuLysSerValArgSerLysValGlySerThrGluA 917

    1307 ATATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTT 1356
         ::|||::::::      |||||||||
     917 snIleLysHisGlnProGlyGlyGly........................ 925

    1357 GATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAAC 1406
                     |||   :::|||::::::|||   :::      |||||
     926 ...........ArgAlaLysValGluLysLysThrGluAlaAlaThrTh 938

    1407 TCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGA 1456
         |     :::   |||   |||:::                |||     :
     938 rAlaGlyLysProGluProAsnAlaValThrLysAlaAlaGlySerIleA 955

    1457 GTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGAAAATAAAACAGGG 1506
         ::|||      |||      |||      :::         :::
     955 laSerAlaGlnLysProProAlaGlyLysValGlnIleValSerLysLys 971

    1507 TTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGT 1556
         ::::::::::::|||:::   :::      ::::::|||   :::
     972 ValSerTyrSerHisIleGlnSerLysCysValSerLysAspAsnIle.. 987

    1557 AACAGGAGGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAA 1606

     987 .................................................. 987

    1607 CCTCGGCACCTGATAAACAT......GGGGTTTATCAAGCGACAGTGGAA 1650
                          ||||||      |||      ::::::   ::::::
     988 .............LysHisValProGlyCysGlyAsnValGlnIleGln 999

    1651 ATTAAAAAGCCTGATGGAAGTTGGGAGGTGAAAACGAAAAAAGGTGGGAA 1700
             ||||||   |||   |||      :::|||:::::::|||   |||:::||
```

```
   1000 AsnLysLysValAspIleSer...LysValSerSerLysCysGlySerLy 1015

   1701 AGTGATGACCAAGCACACCATGTTCCCAAAAGATTGGGATGAGGCTAGAA 1750
          |:::        ||||||                 |||
   1015 sAlaAsnIleLysHisLysProGlyGlyGlyAspValLysIleGluSerG 1032

   1751 TTAGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGAT 1800
           :::   :::      :::       :::::::::::         |||
   1032 lnLysLeuAsnPheLysGluLysAlaGlnAlaLysValGlySerLeuAsp 1048

   1801 AATAAATGGCAGGGTACAAGTAAATCGGGTATTAAAATAGAAGGA 1845
          |||       :::      :::    :::::::::|||     ||||||
   1049 AsnValGlyHisPheProAlaGlyGlyAlaValLysThrGluGly 1063

seq_name: SwissProt_40:YWR1_CAEEL

seq_documentation_block:
ID   YWR1_CAEEL     STANDARD;      PRT;  1223 AA.
AC   Q10925;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative tyrosine-protein kinase B0302.1 in chromosome X
DE   (EC 2.7.1.112).
GN   B0302.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Du Z.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC       DOMAIN. HIGHEST TO THE SYK/ZAP-70 SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U41032; AAA82367.1; -.
DR   HSSP; P11362; 1AGW.
DR   WormPep; B0302.1; CE03866.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein; Transferase; Tyrosine-protein kinase;
KW   ATP-binding.
FT   DOMAIN       107    369       PROTEIN KINASE.
FT   NP_BIND      113    121       ATP (BY SIMILARITY).
FT   BINDING      139    139       ATP (BY SIMILARITY).
FT   ACT_SITE     230    230       BY SIMILARITY.
SQ   SEQUENCE   1223 AA;  132940 MW;  19EFFD8E0A815227 CRC64;

alignment_scores:
          Quality:  137.50             Length:     523
            Ratio:  0.593                Gaps:      24

Percent Similarity:  44.359   Percent Identity: 22.753

alignment_block:
 US-09-303-518D-463 x YWR1_CAEEL   ..

 Align seg 1/1  to: YWR1_CAEEL  from: 1  to: 1223

     12 CCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATGC 61
         |||   |||||||||        |||    |||    :::    ||||||    ||
    715 ProAlaAsnIleProCysLeuValProThrProAlaProProAlaProAl 731

     62 ATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTC 111
         |   |||       |||                 |||:::|||::::::
    731 a..HisPheSerGln..............ProValSerSerGlnArgVa 742

    112 GACCGTCAGCATTTCGAAC............................... 130
          :::   :::    :::|||
    742 lAlaGlnGlnGlnGlnAsnThrLeuGlnLysAlaLeuAsnAspGluLeuL 759

    131 ...................CCGACGGGAAATACCACCTATTCGGCAGCA 160
                           |||||||||:::|||:::        :::
    759 ysGlyAsnLeuAsnLysArgProThrGlyThrThrAlaProProSerAsn 775

    161 GGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAA 210
         |||    :::    |||    |||    :::
    776 GlyPheAsnAlaProArgAlaAspValAlaPro................. 786

    211 AGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAA 260
                                  ::::::||||||:::::::::
    787 .......................ValGlnGlnArgProIleSerSerA 795

    261 TATCG...GCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCC... 304
          |||        :::|||:::       ||||||         ::::::
    795 laSerIleProAlaLeuGlnProGlnProIleGlnHisIleGlnLysPro 811

    305 ATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGT 354
         |||:::|||:::    :::       :::||||||      |||    |||||
    812 IleGlnProGlnGlnVal......ArgIleProProSerThrAlaProVa 826

    355 AGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGA 404
         |              ::::::        :::|||      |||    :::|
    826 l...........GlnLysProValGlnValSerAlaProThrHisSerA 839

    405 ACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCG 454
         ||:::    |||    |||:::           :::          |||        |||
    839 snValAlaProThrThrSerSerGlnAlaSerAlaAspAlaArgAsnPro 855

    455 CTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCC 504
         ||||||                                    |||:::    ||
    856 LeuProPro............................LysThrSerPr 862

    505 CAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCT 554
         |    :::|||    :::        |||    |||:::||||||     ::::::
    862 oProValSerAsnThr......ProIleThrValAlaProValHisAlaA 877

    555 TGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCG 604
           ||||||:::|||
    877 laProThrThrSer.................................. 881

    605 ACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGC 654
                                   |||||||||     :::          ::
    882 .........................AlaProSerThrSerValValTh 889

    655 AATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCAT 704
         :         |||        :::||||||    |||:::|||::::::
    889 rArgArgProThrSerThrThrAla...GlnMetSerAspGluGluArgA 905

    705 CGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAA 754
           :::|||                    ::::::|||:::|||
    905 rgSerArgIleAlaMetAspIleSerSerAlaLeuPro............ 917

    755 GCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACC 804
         |||    :::         |||     :::     ::::::           :::|||
```

```
 918 AlaProSerAlaLeuLeuTyrGlySerAsnSerThrSerSerLeuProSe 934

 805 GAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCA. 853
         :::              ||||||:::::   :::     |||:::::
 934 rAlaAlaValSerThrAlaSerSerValProSerThrAlaArgAspAsnP 951

 854 .....AAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACC 898
          :::|||   |||:::||||||    :::              |||
 951 roValGluThrArgProSerGlnProHisVal..............Thr 962

 899 CCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT...TTATGGCA 945
          |||   |||:::   :::|||              |||   :::   ||
 963 MetProProLysLysSerSerGluProIleLeuSerSerGluValLeuGl 979

 946 GCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGG 995
     ||||:::            :::         |||      :::   :::
 979 nProThrArgLeuProSerAlaThrThrSerGlnAlaLysProValThrG 996

 996 CGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCAT 1045
               ||||||
 996 lnProIleArgHis.................................. 1000

1046 TGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT.GCGGCATA 1094
               :::||||||   ||||||::::::|||   |||:::
1001 .........ProSerProProValAlaThrValIleProThrAlaValVa 1014

1095 CGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGG 1144
         |||   |||               |||:::|||      |||||||:::
1014 lAspLysLysProVal.........SerGlnAsnGlnGlySerAsnValP 1028

1145 AGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACC......GTGCCG 1188
                          |||||||||:::|||:::         |||
1028 ro...........LeuPheAsnIleThrAsnSerSerAsnGlyTyrPro 1040

1189 CCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGAC 1238
          ||||||                              :::|||:::
1041 GlnLeuAsnGly..........................TyrProAsnTy 1048

1239 AGGCGTACCGTTTGACGGT.......................AAAGGGT 1264
       |||      |||:::::                        :::|||:
1048 rGlyAsnGlyPheGlnAlaTyrGlyTyrGlyMetAsnTyrHisGlnGlyT 1065

1265 TTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAA 1314
     ::|||:::::::::    :::    :::|||    :::    :::
1065 yrProGlyTyrGlnGlyTyrAsnSerTyrGlyAsnGlyMetGlyGlnLeu 1081

1315 GAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAA 1364
        |||:::   :::::::::   :::      |||:::      :::::
1082 AlaLeuThrHisAsnAlaValThrSerLeuProProLeuValProSerGl 1098

1365 ACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGC 1414
     :   |||::::::                       |||::::::
1098 uAsnArgPheSerGly....................ThrAlaGlnProL 1108

1415 AGGTGGAGAAAAATGTTCAGGAA.........ACGAGAAGAAGGAGTCAG 1455
           |||::::::::::::|||        |||::::::|||:::
1108 euGlyGluSerAspIleMetGluPheLeuGlyThrGlnGlnArgGlnAla 1124

1456 AGTAGTCAGTTTAAAGCC 1473
     :::|||:::   :::|||
1125 GlySerSerSerArgAla 1130
```

seq_name: SwissProt_40:VRP1_YEAST

```
seq_documentation_block:
ID   VRP1_YEAST     STANDARD;      PRT;   817 AA.
AC   P37370; Q06133;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Verprolin.
GN   VRP1 OR MDP2 OR END5 OR YLR337W OR L8300.13.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A364;
RX   MEDLINE=95058201; PubMed=7968536;
RA   Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E.;
RT   "A proline-rich protein, verprolin, involved in cytoskeletal
RT   organization and cellular growth in the yeast Saccharomyces
RT   cerevisiae.";
RL   Mol. Microbiol. 10:585-596(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA   Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA   Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA   Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA   Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA   Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA   Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN CYTOSKELETAL ORGANIZATION AND CELLULAR
CC       GROWTH. MAY EXERT ITS EFFECTS ON THE CYTOSKELETON DIRECTLY, OR
CC       INDIRECTLY VIA PROLINE-BINDING PROTEINS (E.G. PROFILIN) OR
CC       PROTEINS POSSESSING SH3 DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z26645; CAA81388.1; -.
DR   EMBL; U19028; AAB67263.1; -.
DR   PIR; S39626; S39626.
DR   SGD; S0004329; VRP1.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF02205; WH2; 2.
DR   SMART; SM00246; WH2; 2.
KW   Cytoskeleton; Repeat.
FT   DOMAIN        5     14       POLY-PRO.
FT   DOMAIN      239    245       POLY-PRO.
FT   DOMAIN      349    357       POLY-PRO.
FT   DOMAIN      396    406       POLY-PRO.
FT   DOMAIN      424    431       POLY-PRO.
FT   DOMAIN      462    468       POLY-SER.
FT   DOMAIN      704    708       POLY-PRO.
FT   CONFLICT    308    308       P -> R (IN REF. 1).
FT   CONFLICT    350    350       A -> R (IN REF. 1).
FT   CONFLICT    689    689       V -> E (IN REF. 1).
FT   CONFLICT    710    817       PSTMDTGTSNSPSKNLKQRLFSTGGSTLQHKHNTHTNQPDV
FT                                DVGRYTIGGSNSIVGAKSGNERIVIDDSRFKWTNVSQMPKP
FT                                RPFQNKTKLYPSGKGSSVPLDLTLFT -> HLRWIPVPLIA
FT                                PVKTLNNGYFLQVDRRCNTSIIRIQINQMLM (IN REF.
FT                                1).
SQ   SEQUENCE   817 AA;  82593 MW;  24C7522D5B1CA1C8 CRC64;
```

```
alignment_scores:
           Quality:  137.00           Length:    372
             Ratio:  0.878              Gaps:     18
 Percent Similarity:  41.935   Percent Identity: 23.656

alignment_block:
 US-09-303-518D-463 x VRP1_YEAST   ..

 Align seg 1/1  to: VRP1_YEAST  from: 1  to: 817

     275 TCCGCTTTTCCGATCACGGGCACAAATTCCATTCGC...CCTTCGACAAC 321
```

```
        ||||||          :::||||||  |||   :::  ||||||  ::
  64 SerAlaSerGlySerSerGlyThrValSerSerLysGlyProSerMetSe 80

 322 CATGCCTCACATTCCGATTCTGACGAAGCCGGTA................ 355
      :  |||  ||||||   :::      |||
  80 rAlaProProIleProGlyMetGlyAlaProGlnLeuGlyAspIleLeuA 97

 356 ........GTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACG 397
                :::|||                       |||:::::
  97 laGlyGlyIleProLys...............LeuLysHisIleAsnAsn 108

 398 GATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGC 447
     :::::::::|||   |||        :::           |||:::||
 109 AsnAlaSerThrLysProSerProSer...............AlaSerAl 120

 448 TATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGG 497
     |   |||:::|||                                     |
 120 aProProIlePro................................GlyA 126

 498 CGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGAC 547
     ||:::|||:::::::::|||             |||        ||||||
 126 laValProSerValAlaAlaPro......ProIleProAsnAlaProLeu 140

 548 AACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGA 597
     :::         |||:::|||   ::::::|||:::
 141 SerProAlaProAlaValProSerIleProSer................. 151

 598 GTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAG 647
                   |||::::|||||||||        |||:::      :
 152 .............SerSerAlaProPro......IleProAspIleProS 162

 648 ATCGGGCAATGCCGCCG..................AAGCCTTCAACGGCA 679
     ::    |||   ||||||                   :::|||:::
 162 erSerAlaAlaProProIleProIleValProSerSerProAlaProPro 178

 680 CTGCAGATATCGTCAAAAACATCA.......................... 703
     |||    :::|||:::    ::::::
 179 LeuProLeuSerGlyAlaSerAlaProLysValProGlnAsnArgProHi 195

 704 ........................TCGGCGCGGCAGGAGAAATTGTCGG 728
                              ::::::|||:::|||    |||
 195 sMetProSerValArgProAlaHisArgSerHisGlnArgLysSerSerA 212

 729 CGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCA 778
            ::::::|||
 212 snIleSerLeuProSer................................ 217

 779 TGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAAC 828
                 |||       ||||||
 218 ............ValSerAlaProProLeuPro................ 224

 829 GATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCG 878
                           |||   :::::::|||   :::   |||
 225 ......................SerAlaSerLeuProThrHisValSerA 234

 879 CGATTGGGCAGTCCAAAACCCCAATGCCG..................... 907
              |||:::      |||  |||
 234 snProProGlnAlaProProProProProThrProThrIleGlyLeuAsp 250

 908 CACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCC..... 952
        |||      |||||||:::        ::::::       ||||||
 251 SerLysAsnIleLysProThrAspAsnAlaValSerProProSerSerGl 267

 953 ....CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGG 998
          |||:::   ||||||   :::   :::|||   :::|||      :
 267 uValProAlaGlyGlyLeuProPheLeuAlaGluIleAsnAlaArgArgS 284

 999 CATCACGGCACATC.........CTGTCAAGCGGTCGCAGATGGGCGCGA 1039
     ::::::|||                 :::||||||   :::       ::::::
 284 erGluArgGlyAlaValGluGlyValSerSerThrLysIleGlnThrGlu 300

1040 TCGCATTGCCGAAAGGGAAATCCGCCG.........TCAGCGACAATTTT 1080
     :::|||           ::::::||||||            ||||||      ::
 301 AsnHisLysSerProSerGlnProProLeuProSerSerAlaProProIl 317

1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATAT 1130
      :|||       |||:::|||          |||   :::   |||     |
 317 eProThrSerHisAlaProProLeuProProThrAlaProProProProS 334

1131 CCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCA......... 1171
      ||:::   :::   :::|||:::        |||   ||||||
 334 erLeuProAsnValThrSerAlaProLysLysAlaThrSerAlaProAla 350

1172 ...CCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCA 1218
         ||||||   |||                   |||   ||||||:::   ::
 351 ProProProProProLeu.........ProAlaAlaMetSerSerAlaSe 364

1219 GACCAACGCCACCCGA 1234
      :||||||:::::::::
 364 rThrAsnSerValLys 369
```

seq_name: SwissProt_40:VGLX_HSVEB

```
seq_documentation_block:
ID   VGLX_HSVEB     STANDARD;      PRT;   797 AA.
AC   P28968;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   Glycoprotein X precursor.
GN   71.
OS   Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=31520;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295566; PubMed=1318606;
RA   Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT   "The DNA sequence of equine herpesvirus-1.";
RL   Virology 189:304-316(1992).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M86664; AAB02506.1; -.
DR   PIR; H36802; VGBEX1.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    797       GLYCOPROTEIN X.
FT   DOMAIN       23    465       SER/THR-RICH.
FT   TRANSMEM    766    790       POTENTIAL.
FT   CARBOHYD    590    590       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   797 AA;  80342 MW;  50C9ED9211F5E5B2 CRC64;

alignment_scores:
            Quality:  136.50            Length:    569
              Ratio:  0.559               Gaps:     20
 Percent Similarity:  42.882   Percent Identity: 21.968

alignment_block:
 US-09-303-518D-463 x VGLX_HSVEB   ..

 Align seg 1/1  to: VGLX_HSVEB  from: 1  to: 797

 110 TCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGC 159
      ||||||::::::   ::::::   |||::::::::::::   |||::::::
  22 SerThrThrThrThrGluThrThrThrSerSerSerSerThrSerGlySe 38

 160 AGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACA 209
```

```
          :|||   |||      |||:::|||                 |||    :
  38 rGlyGlnSerThrSerSerGlyThr.................ThrAsnS 49

 210 AAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAA 259
     :::::   |||   :::|||
  49 erSerSerSerProThrThr............................. 55

 260 ATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCG 309
                  |||   |||:::   |||   |||:::|||:::::
  56 SerProProThrThrSerSerSerProProThrSerThrHisThrSerSe 72

 310 CCCTTCG......ACAACCATGCCTCACATTCCGATTCTGACGAAGCCGG 353
      ||||||      :::|||:::                 |||:::
  72 rProSerSerThrSerThrGlnSerSerSerThrAlaAlaThrSerSerS 89

 354 TAGTCCCGTTGACGGATTCAGCCTTTACCGCA.........TCCATTGGG 394
       :::|||   |||   |||:::   |||:::         |||   :::
  89 erAlaProSerThrAlaSerSerThrThrSerIleProThrSerThrSer 105

 395 ACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGC 444
     |||:::|||:::|||       |||||||||   |||         :::::
 106 ThrGluThrThrThrThrThrProThrAlaSerThrThrThrProThrTh 122

 445 GGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAA 494
     :::          |||   :::   :::     ||||||:::|||
 122 rThrThrAlaAlaProThrThrAlaAlaThrThrThrAlaValThrThrA 139

 495 AGGCG............................................ 499
       |||
 139 laAlaSerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSer 155

 500 .............................TTGCCCAAAATATCCGCCTC 519
                                         ::::::::
 156 ThrProThrThrThrThrProThrSerThrThrThrThrThrAlaThrTh 172

 520 AACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCA 569
     :|||  ||||||||||||:::      :::            |||::::::|
 172 rThrValProThrThrAlaSerThrThrThrAspThrThrThrAlaAlaT 189

 570 CAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCG 619
     ||       |||                 |||:::   ::::::|||
 189 hrThrThrAla.........AlaThrThrThrAlaAlaThrThrThrAla 202

 620 CCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCC 669
                 ||||||   :::   |||        :::
 203 AlaThrThrThrAlaAlaThrThrThrAlaAlaThrThrThrAlaAlaTh 219

 670 TTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGA 719
       :::||||||          ||||||   |||::::::|||
 219 rThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrThrA 236

 720 AATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACA 769
          ::::::   |||                  |||   :::   |||
 236 laAlaThrThrThrAlaAlaThrThrThr......AlaAlaThrThrThr 250

 770 TTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCG 819
           :::   ||||||   :::                |||
 251 AlaAlaThrThrThrAlaAlaThrThrThrGlySerProThrSerGlySe 267

 820 CGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAG. 868
       :::||||||                 |||   :::   |||:::
 267 rThrSerThrThrGlyAla...........SerThrSerThrProSerA 280

 869 .................................CAGCCATCCGCGATT 883
                                         :::   ||||||
 280 laSerThrAlaThrSerAlaThrProThrSerThrSerThrSerAlaAla 296

 884 GGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAAT 933
     :::   ::::::||||||   |||   :::|||         ||||||
 297 AlaThrThrSerThrProThrProThrSerAla...AlaThrSerAlaGl 312

 934 ATCTTTATGGCAG.................................C 947
       |||      :::                                 |
 312 uSerThrThrGluAlaProThrSerThrProThrThrAspThrThrThrP 329

 948 CATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG 997
      |||||    :::              |||   |||:::|||:::   :::
 329 roSerGluAlaThrThrAlaThrThrSerProGluSerThrThrValSer 345

 998 GCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTG 1047
      ||||||              ::::::               ::::::||||||
 346 AlaSerThrThrSerAlaThrThrThrAlaPheThrThrGluSerHisTh 362

1048 CCGAAAGGGAAAT.......................CCGCCGTCAGCG. 1072
           :::                              ||||||:::
 362 rSerProAspSerSerThrGlySerThrSerThrAlaGluProSerSerT 379

1073 .....ACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACC 1117
          |||:::   |||         |||||||:::|||      :::
 379 hrPheThrLeuThrProSerThrAlaThrProSerThrAspGlnPhe... 394

1118 ATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAAC 1167
                          |||   ||||||:::::::::::::::
 395 ....................ThrGlySerSerAlaSerThrGluSerAs 404

1168 ATCACC....TCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAAC 1213
       |||      ||||||||||||||   ::::::   :::::::::
 404 pSerThrAspSerSerThrValProThrThrGlyThrGluSerIle.... 419

1214 TGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGG 1263
                :::       :::|||           :::   |||
 420 ........ThrGluSerSerSerThrThrGluAlaSerThrAsnLeuGly 433

1264 TTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCA 1313
           ::::::|||:::            :::|||      |||    ::
 434 SerSerThrTyrGluSerThrGluAlaLeuGluThrPro...AspGlyAs 449

1314 AGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGA 1363
     :        ||||||:::       |||:::   :::|||
 449 nThrThrSerGlyAsnThrThrProSerProSerPro............. 461

1364 AACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAG 1413
                             :::      :::::::::      :::
 462 ......................ArgThrProSerPheAlaAspThrGln 470

1414 CAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCA 1463
     |||:::      |||      :::|||::::::   :::   ::::::::
 471 GlnThrProAspAsnGlyValSerThrGlnHisThrThrIleAsnAspHi 487

1464 GTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATT 1513
     :        |||:::||||||::::::
 487 sThrThrAlaAsnAlaGlnLysHis........................A 496

1514 TTAATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGA 1563
        :::|||      |||         :::::::::|||:::      |||
 496 laGlyHisHisArgGlyArgAlaGlyGlyArgArgGlySerProGlnGly 512

1564 GGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGC 1613
      |||        |||         |||:::           |||    ::
 513 GlySerHisThrThrProHisProAspArgLeu.........ThrProSe 526

1614 ACCTGAT 1620
      :||||||
 526 rProAsp 528
```

seq_name: SwissProt_40:AMYB_PAEPO

seq_documentation_block:

```
ID   AMYB_PAEPO     STANDARD;      PRT;  1196 AA.
AC   P21543;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Beta/alpha-amylase precursor [Includes: Beta-amylase (EC 3.2.1.2);
DE   Alpha-amylase (EC 3.2.1.1)].
OS   Paenibacillus polymyxa (Bacillus polymyxa).
```

```
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Paenibacillus.
OX   NCBI_TaxID=1406;
RN   [1]
RP   SEQUENCE OF 1-936 FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=72;
RX   MEDLINE=87165765; PubMed=2435707;
RA   Kawazu T., Nakanishi Y., Uozumi N., Sasaki T., Yamagata H.,
RA   Tsukagoshi N., Udaka S.;
RT   "Cloning and nucleotide sequence of the gene coding for enzymatically
RT   active fragments of the Bacillus polymyxa beta-amylase.";
RL   J. Bacteriol. 169:1564-1570(1987).
RN   [2]
RP   SEQUENCE OF 689-1196 FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=72;
RX   MEDLINE=89123046; PubMed=2464578;
RA   Uozumi N., Sakurai K., Sasaki T., Takekawa S., Yamagata H.,
RA   Tsukagoshi N., Udaka S.;
RT   "A single gene directs synthesis of a precursor protein with beta-
RT   and alpha-amylase activities in Bacillus polymyxa.";
RL   J. Bacteriol. 171:375-382(1989).
RN   [3]
RP   SEQUENCE OF 1-776 FROM N.A.
RC   STRAIN=ATCC 8523;
RX   MEDLINE=87231094; PubMed=2438660;
RA   Rhodes C., Strasser J., Friedberg F.;
RT   "Sequence of an active fragment of B. polymyxa beta amylase.";
RL   Nucleic Acids Res. 15:3934-3934(1987).
RN   [4]
RP   DISULFIDE BOND, AND MUTAGENESIS OF CYSTEINE RESIDUES.
RX   MEDLINE=91215008; PubMed=1827035;
RA   Uozumi N., Matsuda T., Tsukagoshi N., Udaka S.;
RT   "Structural and functional roles of cysteine residues of Bacillus
RT   polymyxa beta-amylase.";
RL   Biochemistry 30:4594-4599(1991).
CC   -!- FUNCTION: THE PRECURSOR PROTEIN IS PROTEOLITICALLY CLEAVED TO
CC       PRODUCE MULTIFORM BETA-AMYLASES AND A 48 kDa ALPHA-AMYLASE AFTER
CC       SECRETION.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC       polysaccharides so as to remove successive maltose units from the
CC       non-reducing ends of the chains.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC       linkages in oligosaccharides and polysaccharides.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO FAMILY 14 OF
CC       GLYCOSYL HYDROLASES (BETA-AMYLASES).
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 13 OF
CC       GLYCOSYL HYDROLASES, ALSO KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M15817; AAA85446.1; -.
DR   EMBL; Y00150; CAA68344.1; -.
DR   PIR; A29130; A29130.
DR   PIR; A29108; A29108.
DR   PIR; A32251; A32251.
DR   HSSP; P36924; 1B9Z.
DR   InterPro; IPR000461; Alpha_amylase.
DR   InterPro; IPR001554; Glyco_hydro_14.
DR   Pfam; PF00128; alpha-amylase; 1.
DR   Pfam; PF02806; alpha-amylase_C; 1.
DR   Pfam; PF01373; Glyco_hydro_14; 1.
DR   PRINTS; PR00750; BETAAMYLASE.
DR   PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR   PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW   Multifunctional enzyme; Hydrolase; Glycosidase; Signal;
KW   Polysaccharide degradation; Repeat.
FT   SIGNAL        1     35
FT   CHAIN        36   1196       BETA/ALPHA-AMYLASE.
FT   DOMAIN       36    454       BETA-AMYLASE.
FT   REPEAT      455    558
FT   REPEAT      565    668
FT   DOMAIN      669   1196       ALPHA-AMYLASE.
FT   DISULFID    118    126
FT   ACT_SITE    124    124       BY SIMILARITY.
FT   ACT_SITE    198    198       BY SIMILARITY.
FT   MUTAGEN     118    118       C->S: 5 FOLD DECREASE IN ACTIVITY.
FT   MUTAGEN     126    126       C->V: 20 FOLD DECREASE IN ACTIVITY.
FT   MUTAGEN     358    358       C->S: 60 FOLD DECREASE IN ACTIVITY.
FT   CONFLICT      1      1       M -> MIGL (IN REF. 3).
FT   CONFLICT     67     67       N -> S (IN REF. 3).
FT   CONFLICT    100    100       N -> D (IN REF. 3).
FT   CONFLICT    154    154       S -> N (IN REF. 3).
FT   CONFLICT    177    177       E -> Q (IN REF. 3).
FT   CONFLICT    227    228       NA -> KS (IN REF. 3).
FT   CONFLICT    330    330       G -> S (IN REF. 3).
FT   CONFLICT    425    425       N -> S (IN REF. 3).
FT   CONFLICT    493    493       D -> A (IN REF. 3).
FT   CONFLICT    532    532       S -> L (IN REF. 3).
FT   CONFLICT    559    559       A -> T (IN REF. 3).
FT   CONFLICT    665    665       A -> T (IN REF. 3).
FT   CONFLICT    681    681       D -> N (IN REF. 3).
FT   CONFLICT    686    686       T -> A (IN REF. 3).
FT   CONFLICT    725    728       AFTS -> VFSP (IN REF. 3).
FT   CONFLICT    736    736       N -> K (IN REF. 3).
FT   CONFLICT    741    741       N -> S (IN REF. 3).
FT   CONFLICT    758    758       S -> N (IN REF. 3).
SQ   SEQUENCE   1196 AA;  130893 MW;  A41EA6B70F257064 CRC64;

alignment_scores:
            Quality:  133.00            Length:    717
              Ratio:  0.426               Gaps:     38
 Percent Similarity:  43.515   Percent Identity: 21.060

alignment_block:
 US-09-303-518D-463 x AMYB_PAEPO   ..

 Align seg 1/1  to: AMYB_PAEPO  from: 1  to: 1196

      71 CCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAG 120
         |||:::|||:::            |||||||||
     187 ProLysIleTyr...........LeuSerGly................ 193

     121 CATTTCGAACCCGACGGGAAAT...........ACCACCTATTCGGCAG 158
                :::|||:::|||:::            :::   |||   ||||
     194 .......GlyProSerGlyGluLeuArgTyrProSerTyrTyrProAlaA 208

     159 CAGGGGGGAGCTTGCCNAGC............................. 178
         |||||   |||   |||:::
     208 laGlyTrpSerTyrProGlyArgGlyLysPheGlnAlaTyrThrGluThr 224

     179 ..............GCAACGGCCATATCGGAT................ 196
                           ||||||::::::|||
     225 AlaLysAsnAlaPheArgThrAlaMetAsnAspLysTyrGlySerLeuAs 241

     197 ...............TGGGAAACATACAAAGCCATCA.........GTT 221
                         |||   |||            |||     :::
     241 pLysIleAsnAlaAlaTrpGlyThr.LysLeuThrSerLeuSerGlnIle 257

     222 GGGCCACCTGATGAT..............TCAACAGGCGGCCGTTGAAG 256
         :::||||||   |||              ::::::||||||
     258 AsnProProThrAspGlyAspGlyPheTyrThrAsnGlyGlyTyrAsnSe 274

     257 GAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCAT 306
            ||| |||              :::|||              ::::::
     274 rAlaTyr.Gly..............LysAspPheLeuSerTrpTyrGln 285

     307 TCGCCCTTCGACAACCAT.................GCCTCACATTCCGA 338
         |||   :::::::::|||                 |||:::|||::::
```

```
 286 SerValLeuGluLysHisLeuGlyValIleGlyAlaAlaAlaHisLysAs 302

 339 TTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCC 388
     :  |||::::::   :::        |||   ::::::     :::|
 302 nPheAspSerValPheGlyValArgIleGlyAlaLysIleSerGlyLeuH 319

 389 ATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAG 438
     |||||:::           ::::::||||||       :::        |||
 319 isTrpGln......MetAsnAsnProAlaMetProHisGlyThrGluGln 333

 439 GGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGA 488
     :::|||||||||                          |||:::|||::
 334 AlaGlyGlyTyr..........................TyrAspTyrAs 341

 489 CATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCA 538
     :
 341 n................................................. 341

 539 GCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTG 588
               |||||    ::::::|||   :::|||        |||
 342 ..........ArgLeuIleGlnLysPheLysAspAlaAspLeuAspLeu 354

 589 ACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGA 638
                                     |||          ||
 355 ................................ThrPheThrCysLeuGl 360

 639 GCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC......AACGGCACTG 682
     |::::::    ||||||:::|||     :::          :::      :
 360 uMetSerAspSerGlyThrAlaProAsnTyrSerLeuProSerThrLeuV 377

 683 CAGATATCGTCAAAAACATCATCGGCGCGGCAGGA............... 717
     ::|||    |||::::::|||    :::|||    |||
 377 alAspThrValSerSerIleAlaAsnAlaLysGlyValArgLeuAsnGly 393

 718 ......GAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGG 761
              :::       :::|||::::::   |||    |||:::|||
 394 GluAsnAlaLeuProThrGlyGlySerGlyPheGlnLysIleGluGluLy 410

 762 CTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACA 811
        :::                |||||||:::   ||||||
 410 sIleThrLysPheGlyTyrHisGlyPheThrLeuLeu............ 422

 812 AGATGGCGCGCATCAACGATTTG.......................GCA 837
            ||||||||:::|||                           :::
 423 ........ArgIleAsnAsnLeuValAsnAsnAspGlySerProThrGly 436

 838 GATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGC 887
     ::::::::    :::|||:::|||         :::       :::
 437 GluLeuSerGlyPheLysGlnTyrIleIleSerLysAlaLys........ 450

 888 AGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT 937
                |||:::         |||              :::|||:
 451 ..........ProAspAsnAsnGlyGlyThrGlyAsnLysValThrIleT 464

 938 TTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT......GTCCGGGGA 981
     ::                        ||||||:::::::::       ::::::
 464 yrTyrLys............LysGlyPheAsnSerProTyrIleHisTyr 476

 982 AAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGAT 1031
     :::   :::    |||:::    ||||||       :::      |||
 477 ArgProAlaGlyGlySerTrpThrAlaAlaProGlyValLysMetGlnAs 493

1032 GGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTG 1081
       :::    |||:::                |||   :::   ::::::   :::::::
 493 pAlaGluIleSer.........GlyTyrAlaLysIleThrValAspIleG 507

1082 CCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATC 1131
     :::::|||:::       :::                      ::::::|||
 507 lySerAlaSerGlnLeuGluAlaAlaPheAsnAspGlyAsnAsnAsnTrp 523

1132 CGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAAC 1181
        ||||||       ::::::|||:::   ::::::   |||    |||||
 524 AspSerAsnAsnThrLysAsnTyrSerPheSerThrGlyThr...SerTh 539

1182 CGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACC 1231
     |         |||:::|||:::   |||:::                :::      |
 539 rTyrThrProGlyAsnSerGlyAsnAlaGlyThrIleThrSerGlyAlaP 556

1232 CGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT......... 1272
     ||          :::    |||    ||||||    |||        |||
 556 roAlaGlyAlaAsnProGlyAspGlyGlyGlyThrThrAsnLysValThr 572

1273 ......TTTGAGAAG..............CACGTGAAATAT........ 1293
            ::::::|||                   ::::::    |||
 573 ValTyrTyrLysLysGlyPheAsnSerProTyrIleHisTyrArgProAl 589

1294 ........................GATACGAAGCTCGATATTCAAGAAT 1318
                              :::|||::::::         |||:
 589 aGlyGlySerTrpThrAlaAlaProGlyValLysMetGlnAspAlaGluI 606

1319 TATCGGGG......................GGCGGTATACCTAAGGCT 1344
     ::||||||                      |||:::         :::
 606 leSerGlyTyrAlaLysIleThrValAspIleGlySerAlaSerGlnLeu 622

1345 AAGCCTGTGTTT...GATGCGAAACCGAGATGGGAGGTTGATAGGAAGCT 1391
     :::    :::|||     |||::::::    :::|||:::    ::::::
 623 GluAlaAlaPheAsnAspGlyAsnAsnAsnTrpAspSerAsnAsnThrLy 639

1392 TAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGA 1441
      |||    |||    :::                              |||
 639 sAsnTyrLeuPheSer.......................ThrGlyThrS 648

1442 GAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGG 1491
           :::    :::|||:::         |||    :::    |||
 648 erThrTyrThrProGlySerAsnGlyAlaAlaGlyThrIleArg...... 662

1492 GAAAATAAAACAGGGTTAGATTTTAATCATTTTATA.............. 1527
                ||||||         :::        :::
 663 .........ThrGlyAlaProSerGlySerValLeuSerValValThrSe 676

1528 .......GGTGGTGATATCAATAAGAAAGGCACAGTAACAGGA....... 1563
               :::    |||:::|||:::          |||||||||
 676 rThrTyrAlaThrAspLeuAsnGlu.........ValThrGlyProIleG 690

1564 ..GGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCG 1611
           ::::::|||:::        :::::::::|||      :::|||
 690 lnThrGluLysLeuSerGlyValSerLeuAsnValSerThrSerThrTyr 706

1612 GCACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCC 1661
     ||||||:::::::::||||||    ::::::|||::::::    :::    ||
 707 AlaProAsnSerAsnGlyVal...GluValThrAlaGlnThrGluAlaPr 722

1662 TGATGGAAGTTGGGAGGTGAAAACGAAAAAAGGTGGGAAAGTGATGACCA 1711
     |:::|||::::::              |||:::             :::::::::
 722 oSerGlyAlaPhe.........ThrSerMetAspLeuGlyThrLeuSerA 736

1712 AGCACACCATGTTCCCAAAAGATTGGGATGAGGCTAGAATTAGGGCTGAA 1761
     ::    |||    :::         ||||||:::::         |||
 736 snProThrSerLeuAsnThrAspTrpSerLysGlnSerIleTyrPheIle 752

1762 GTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGATAAT........ 1803
     :::|||:::    :::::::::               :::||||||
 753 MetThrAspArgPheSerAsnGlyAspProSerAsnAspAsnTyrGlyGl 769

1804 ...........................AAATGGCAGGGTACAAGTAAAT 1825
                                  ||||||:::|||    :::    :
 769 yPheAsnSerAsnAsnSerAspGlnArgLysTrpHisGlyGlyAspPheG 786

1826 CGGGT....................ATTAAAATAGAAGGATTTACC 1851
     ::|||                    ||||||      |||||||||
 786 lnGlyIleIleAsnLysLeuAspTyrIleLysAsnMetGlyPheThr 801
```

seq_name: SwissProt_40:DRPL_HUMAN

seq_documentation_block:
ID DRPL_HUMAN STANDARD; PRT; 1185 AA.

```
AC   P54259;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN   DRPLA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum, and Brain;
RX   MEDLINE=95144175; PubMed=7842016;
RA   Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA   Inoue T., Yamada M.;
RT   "Structure and expression of the gene responsible for the triplet
RT   repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";
RL   Nat. Genet. 8:177-182(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96262314; PubMed=8965642;
RA   Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA   Kidwai A.S., Ashworth R.G., Ross C.A.;
RT   "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT   brain.";
RL   Brain Res. Mol. Brain Res. 36:219-226(1996).
RN   [3]
RP   SEQUENCE OF 470-725 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=93315145; PubMed=8325628;
RA   Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT   "Novel triplet repeat containing genes in human brain: cloning,
RT   expression, and length polymorphisms.";
RL   Genomics 16:572-579(1993).
CC   -!- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC       OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE
CC       LIVER, THYMUS AND LEUKOCYTES.
CC   -!- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC       (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC       ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC       EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC       DISEASE.
CC   -!- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC       PALLIDOLUYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC       DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC       NUCLEUS, RUBRUM, GLOGUS PALLIDUS AND LUYS'BODY. CLINICAL FEATURES
CC       ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC       THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC       IN THE FOURTH.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D31840; BAA06626.1; -.
DR   EMBL; U23851; AAB50276.1; -.
DR   EMBL; L10377; -; NOT_ANNOTATED_CDS.
DR   HSSP; P00651; 1LRA.
DR   MIM; 125370; -.
DR   InterPro; IPR002951; Atrophin.
DR   PRINTS; PR01222; ATROPHIN.
KW   Triplet repeat expansion; Polymorphism.
FT   DOMAIN       73     82       SER/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      302    305       POLY-PRO.
FT   DOMAIN      376    382       POLY-SER.
FT   DOMAIN      386    397       POLY-SER.
FT   DOMAIN      442    447       POLY-PRO.
FT   DOMAIN      479    483       POLY-HIS.
FT   DOMAIN      484    497       POLY-GLN.
FT   DOMAIN      504    507       POLY-PRO.
FT   DOMAIN      564    574       POLY-SER.
FT   DOMAIN      704    707       POLY-PRO.
FT   DOMAIN      802    815       ARG/ALA-RICH (MIXED CHARGE).
FT   DOMAIN      816    827       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      925    934       ARG/GLU-RICH (MIXED CHARGE).
FT   CONFLICT     94     94       MISSING (IN REF. 2).
FT   CONFLICT    333    333       Y -> H (IN REF. 2).
FT   CONFLICT    339    339       M -> I (IN REF. 2).
FT   CONFLICT    541    541       P -> T (IN REF. 3).
FT   CONFLICT   1028   1028       G -> A (IN REF. 2).
SQ   SEQUENCE   1185 AA;  124785 MW;  56C306267331C005 CRC64;

alignment_scores:
           Quality:  132.00             Length:    458
             Ratio:  0.714                Gaps:     22
 Percent Similarity:  40.393   Percent Identity: 22.052

alignment_block:
 US-09-303-518D-463 x DRPL_HUMAN   ..

 Align seg 1/1  to: DRPL_HUMAN  from: 1  to: 1185

      65 CACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGAC 114
         |||    |||             |||   |||:::|||       |||
     253 HisProPro.........ProThrThrProIleSerValSerSerSerGl 266

     115 CGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGG 164
           :::|||   :::   ||||||          |||       :::::||||
     266 yAlaSerGlyAlaProProThrLysProProThrThrProValGlyGlyG 283

     165 GGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCC 214
         ||:::||||||||||||             :::::::::    |||:::   ||
     283 lyAsnLeuProSerAlaProProProAlaAsnPhe.ProHisValThrPr 299

     215 ATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATC 264
         |::::::   ||||||      :::     |||||||||::::::    ||
     299 oAsnLeu...ProProProProAlaLeu..ArgProLeuAsnAsnAlaSe 314

     265 GGCTACA.........TTGTCCGCTTTTCCGATCACGGGC.......... 295
         ||||:::          |||:::|||   |||:::   |||
     314 rAlaSerProProGlyLeuGlyAlaGlnProLeuProGlyHisLeuProS 331

     295 .................................................. 295

     331 erProTyrAlaMetGlyGlnGlyMetGlyGlyLeuProProGlyProGlu 347

     296 .......................ACAAATTCCATTCGCCCTTCGACAAC 321
                                 :::|||:::    |||:::::::::
     348 LysGlyProThrLeuAlaProSerProHisSerLeuProProAlaSerSe 364

     322 CATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATT 371
         :              |||    :::       |||          ::::::|
     364 rSerAlaProAlaProProMetArgPheProTyrSerSerSerSerSerS 381

     372 CAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACG 421
         ||:::    :::||||||    :::::::::::::::::
     381 erSerAlaAlaAlaSerSerSerSerSerSerSerSer............ 393

     422 GCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGG 471
                           :::::::|||   |||:::|||    :::
     394 ..................SerSerAlaSerProPheProAlaSer..... 402

     472 GATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAA 521
                                :::|||||||||:::          |||
     403 .....................GlnAlaLeuProSerTyrProHisSerP 412

     522 CCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACA 571
               |||    |||:::           ::::::    |||          |||
     412 heProProProThrSerLeuSerValSerAsnGlnProProLysTyrThr 428
```

```
 572 ATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCC 621
     :::|||:::|||
 429 GlnProSerLeu.................................... 432

 622 ACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTT 671
      |||:::   |||      |||:::::        ||||||   |||
 433 .ProSerGlnAla...ValTrpSerGlnGlyProProProProProProT 448

 672 CAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAA 721
             |||       :::|||
 448 yrGlyArgLeuLeuAlaAsnSer........................... 455

 722 TTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATT 771
        :::|||:::       |||             :::   ||||||:::
 456 ...AsnAlaHisProGlyProPheProProSerThrGlyAlaGlnSer.T 471

 772 GCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCG 821
       :::|||      :::::::::       |||:::    |||:::
 471 hrAlaHisProProValSerThrHisHisHisHisHisGlnGlnGlnGln 487

 822 CATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAG 871
     :::|||:::                  |||:::      ::::::
 488 GlnGlnGln.................GlnGlnGlnGlnGlnGlnHi 498

 872 CCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAA 921
      |||    :::    |||    |||    |||
 498 sHisGlyAsnSerGlyProProProPro...................... 507

 922 GCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGC 971
                           |||:::   ||||||        :::    :::
 508 ..................GlyAlaPheProHisProLeuGluGlyGly 517

 972 TGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCT....... 1014
                          :::    |||||| |||||||||
 518 ....................SerSerHisHis.AlaHisProTyrAlaM 527

1014 .................................................. 1014

 527 etSerProSerLeuGlySerLeuArgProTyrProProGlyProAlaHis 543

1015 ....................GTCAAGCGGTCGCAGATGGGCGCGATCGC 1043
                          |||:::   ||||||   |||         ::
 544 LeuProProProHisSerGlnValSerTyrSerGlnAlaGlyProAsnGl 560

1044 ATTGCCGAAAGGGAAATCCGCC...GTCAGCGACAATTTTGCCGATGCGG 1090
     :   |||   ::::::|||:::        |||:::::::   :::::::::
 560 yProProValSerSerSerSerAsnSerSerSerSerThrSerGlnGlyS 577

1091 CATAC.....................................GCCAAA 1101
     ::|||                                       |||
 577 erTyrProCysSerHisProSerProSerGlnGlyProGlnGlyAlaPro 593

1102 TACCCGTCCCCTTACCATTCCCGAAATA......TCCGTTCAAACTTGGA 1145
     ||||||   |||     :::|||    :::         |||      |||    |
 594 TyrProPhePro.ProValProThrValThrThrSerSerAlaThrLeuS 610

1146 GCAGCGTTACGGCA..........................AAGAAAACA 1168
     |||:::|||    |||                              |||    :::
 610 erThrValIleAlaThrValAlaSerSerProAlaGlyTyrLysThrAla 626

1169 TCACCTCCTCAACCG 1183
     |||||||||    |||
 627 SerProProGlyPro 631
```

seq_name: SwissProt_40:GTFB_STRMU

```
seq_documentation_block:
ID   GTFB_STRMU     STANDARD;      PRT;  1476 AA.
AC   P08987; O69381; O69384; O69387; O69390; O69396;
DT   01-NOV-1988 (Rel. 09, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextransucrase)
DE   (Sucrose 6-glucosyltransferase).
GN   GTFB.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GS-5;
RX   MEDLINE=87308013; PubMed=3040685;
RA   Shiroza T., Ueda S., Kuramitsu H.K.;
RT   "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL   J. Bacteriol. 169:4263-4270(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT4239, MT4245, MT4251, MT4467, AND MT8148;
RX   MEDLINE=98231643; PubMed=9570124;
RA   Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA   Kimura S., Hamada S.;
RT   "Molecular analyses of glucosyltransferase genes among strains of
RT   Streptococcus mutans.";
RL   FEMS Microbiol. Lett. 161:331-336(1998).
CC   -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC       TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC       OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC       AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC   -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
CC       fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DISEASE: DENTAL CARIES.
CC   -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC       1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC       WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC       FORMS OF GLUCANS.
CC   -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC       BINDING PROTEIN FROM S.MUTANS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17361; AAA88588.1; -.
DR   EMBL; D88651; BAA26101.1; -.
DR   EMBL; D88654; BAA26105.1; -.
DR   EMBL; D88657; BAA26109.1; -.
DR   EMBL; D88660; BAA26113.1; -.
DR   EMBL; D89977; BAA26119.1; -.
DR   PIR; B33135; B33135.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR003318; Glyco_hydro_70.
DR   Pfam; PF01473; CW_binding_1; 13.
DR   Pfam; PF02324; Glyco_hydro_70; 1.
KW   Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35   1476       GLUCOSYLTRANSFERASE-I.
FT   DOMAIN       35   1051       CATALYTIC (APPROXIMATE).
FT   DOMAIN     1097   1476       GLUCAN-BINDING (APPROXIMATE).
FT   REPEAT     1097   1130       A REPEAT.
FT   DOMAIN     1161   1470       5 X TANDEM REPEATS.
FT   REPEAT     1161   1210       1.
FT   REPEAT     1225   1275       2.
FT   REPEAT     1290   1340       3.
FT   REPEAT     1355   1405       4.
FT   REPEAT     1420   1470       5.
FT   VARIANT      62     62       S -> T (IN STRAIN MT4239).
FT   VARIANT      65     65       I -> T (IN STRAINS MT4239, MT4245,
FT                                MT4251, MT4467 AND MT8148).
FT   VARIANT      68     68       A -> V (IN STRAIN MT4239).
FT   VARIANT      78     78       Q -> P (IN STRAIN MT4251).
```

```
FT   VARIANT      86     86       S -> I (IN STRAIN MT4239).
FT   VARIANT      89     89       S -> F (IN STRAIN MT4251).
FT   VARIANT     168    168       K -> N (IN STRAIN MT4251).
FT   VARIANT     276    276       D -> S (IN STRAINS MT4239, MT4245 AND
FT                                MT4251).
FT   VARIANT     399    399       N -> R (IN STRAIN MT4239).
FT   VARIANT     474    474       I -> T (IN STRAIN MT4239).
FT   VARIANT     512    512       K -> R (IN STRAIN MT8148).
FT   VARIANT     519    519       F -> Y (IN STRAIN MT8148).
FT   VARIANT     701    701       T -> I (IN STRAIN MT8148).
FT   VARIANT     708    708       A -> V (IN STRAIN MT8148).
FT   VARIANT     938    938       F -> L (IN STRAIN MT8148).
FT   VARIANT     952    957       YGTPVA -> FGKPVE (IN STRAINS MT4245,
FT                                MT4251 AND MT8148).
FT   VARIANT     963    964       NT -> SV (IN STRAINS MT4245, MT4251 AND
FT                                MT8148).
FT   VARIANT     968    970       VDG -> ADS (IN STRAINS MT4245, MT4251 AND
FT                                MT8148).
FT   VARIANT    1086   1086       A -> T (IN STRAIN MT4239).
FT   VARIANT    1158   1158       S -> N (IN STRAIN MT4239).
FT   VARIANT    1163   1163       H -> Y (IN STRAIN MT4251).
FT   VARIANT    1168   1168       E -> K (IN STRAIN MT8148).
FT   VARIANT    1182   1182       Y -> C (IN STRAIN MT8148).
FT   VARIANT    1234   1234       A -> P (IN STRAIN MT4239).
FT   VARIANT    1263   1263       H -> P (IN STRAIN MT8148).
FT   VARIANT    1263   1263       H -> R (IN STRAINS MT4239, MT4245 AND
FT                                MT4251).
FT   VARIANT    1264   1264       H -> Y (IN STRAINS MT4245 AND MT4251).
FT   VARIANT    1272   1272       G -> S (IN STRAINS MT4245 AND MT4251).
FT   VARIANT    1329   1329       Y -> H (IN STRAINS MT4239, MT4245, MT4251
FT                                AND MT8148).
FT   VARIANT    1394   1394       H -> Y (IN STRAINS MT4245 AND MT4251).
FT   VARIANT    1402   1402       G -> S (IN STRAINS MT4245 AND MT4251).
FT   VARIANT    1459   1459       Y -> H (IN STRAIN MT4467).
FT   CONFLICT    570    570       R -> A (IN REF. 1).
FT   CONFLICT    800    817       ADQDVRVAASTAPSTDGK -> LIKMFALRLARPHQQMA
FT                                (IN REF. 1).
FT   CONFLICT   1310   1310       H -> L (IN REF. 1).
SQ   SEQUENCE   1476 AA;  165685 MW;  3479B62B07694D98 CRC64;

alignment_scores:
           Quality:  129.50           Length:    751
             Ratio:  0.415              Gaps:     46
 Percent Similarity:  41.545   Percent Identity:  21.571

alignment_block:
 US-09-303-518D-463 x GTFB_STRMU   ..

 Align seg 1/1   to: GTFB_STRMU   from: 1   to: 1476

      64 GCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGA 113
          ::::::|||        |||:::|||          :::    :::|||
     617 SerTyrAlaLeuLeuLeuThrAsnLysSerSerValProArgValTyrTy 633

     114 CCGTCAGCATTTCGAACCCGACGGGAAATAC.................. 144
              :::    |||        ||||||:::|||
     633 rGlyAspMetPheThrAspAspGlyGlnTyrMetAlaHisLysThrIleA 650

     145 .................CACCTATTCGGCAGCAGG............... 162
                                |||:::   :::|||
     650 snTyrGluAlaIleGluThrLeuLeuLysAlaArgIleLysTyrValSer 666

     163 GGGGAGCTTGCCNAGCGCAACGGCCATATC................... 192
          |||        |||    ||||||    ::::::
     667 GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGluIleIleTh 683

     193 .............GGATTGGGAAACATACAAAGCCATCAGTTGGGCCACC 229
                       |||    |||    :::::::::    :::    |||
     683 rSerValArgTyrGlyLysGlyAlaLeuLysAlaThrAspThrGlyAspA 700

     230 TGATGATTCAACAG.........GCGGCCGTTGAAGGAAATATCGGCTAC 270
                  :::                   |||::::::|||||||||
     700 rgThrThrArgThrSerGlyValAlaValIleGluGlyAsnAsnProSer 716

     271 ATT...GTCCGCTTTTCCGAT..........................CA 290
          :::    ::::::    ||||||                        ||
     717 LeuArgLeuLysAlaSerAspArgValValValAsnMetGlyAlaAlaHi 733

     291 CGGGCACAAATTCCATTCGCCCTTC............GACAAC......C 322
          |    ::::::    :::    |||:::              ||||||
     733 sLysAsnGlnAlaTyrArgProLeuLeuLeuThrThrAspAsnGlyIleL 750

     323 ATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTC 372
             |||    |||||||||::::::    ||||||
     750 ysAlaTyrHisSerAspGlnGluAlaAlaGly.................. 760

     373 AGCCTTTACCGCATCCATTGGACGGATACGAA......CACCATCCCGC 416
             |||    |||              |||       |||         ||
     761 ...LeuValArgTyrThrAsnAspArgGlyGluLeuIlePheThrAlaAl 776

     417 CGAC......GGCTATGACGGGCCACAGGGCGGCGGCTATCCC....... 453
          ||||        ||||||    :::||||||    :::||||||
     776 aAspIleLysGlyTyrAlaAsnProGlnValSerGlyTyrLeuGlyValT 793

     454 ..GCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTT 501
            :::|||    ||||||                            ::::::
     793 rpValProValGlyAla..........................AlaAla 800

     502 GCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACG 551
             |||::::::|||:::          :::          ||||||    ::
     801 AspGlnAspValArgValAlaAlaSerThrAlaProSerThrAspGlyLy 817

     552 GCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAG 601
          :                    |||:::    :::|||:::    ::::::   :::
     817 s.........SerValHisGlnAsnAlaAlaLeuAspSerArgValMetP 831

     602 GCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCG 651
            :::||||||::::::             ::::::    :::    ::::::
     831 heGluGlyPheSerAsnPheGlnAlaPheAlaThrLysLysGluGluTyr 847

     652 GGCAATGCCGCCGAAGCCTTCAAC.....................GGCAC 680
             |||::::::    |||    |||                       |||::
     848 ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluTrpGlyVa 864

     681 TGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 730
          ::::|||                 :::       |||    :::    |||::::
     864 lThrAsp...............PheGluMetAlaProGlnTyrValSerS 876

     731 CAGGCGATGCCGTG.........CAGGGTATAAGCGAAGGCTCAAACATT 771
          ::    |||:::               :::    |||::::::|||        :::
     876 erThrAspGlySerPheLeuAspSerValIleGlnAsnGlyTyrAlaPhe 892

     772 GCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCG 821
          :::        :::    ||||||    :::|||       ||||||    :::
     893 ThrAspArgTyrAspLeuGly...IleSerLysProAsnLysTyrGlyTh 908

     822 CATCAACGATTTGGCAGAT...ATGGCGCAACTCAAAGACTATGCCGCAG 868
             :::||||||:::        :::       |||    :::    :::
     908 rAlaAspAspLeuValLysAlaIleLysAlaLeuHisSerLysGlyIleL 925

     869 CAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATA 918
            ::::::    ||||||:::    ::::::
     925 ysValMetAlaAspTrpValProAspGln..................... 934

     919 GAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGG...AT 965
                                 |||    |||:::|||    |||        ::
     935 .....................MetTyrAlaPheProGluLysGluValVa 944

     966 TGGAGCTGTCCGG...GGAAAATACGGCTTGGGCGGCATCACGGCACATC 1012
          :    |||:::|||        |||||||||                    |
     944 lThrAlaThrArgValAspLysTyrGlyThr..................P 955

    1013 CTGTCAAGCGGTCGCAGATG...GGCGCGATCGCATTGCCGAAAGGGAAA 1059
          |||||        ||||||:::    :::::::::    :::        ||||||
     955 roValAlaGlySerGlnIleLysAsnThrLeuTyrValValAspGlyLys 971
```

```
1060 TCCGCCGTCAGCGACAAT....................TTTGCCGATGC 1088
      |||:::   :::|||:::                   |||   :::
 972 SerSerGlyLysAspGlnGlnAlaLysTyrGlyGlyAlaPheLeuGluGl 988

1089 GGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAA 1138
             ||||||||||||:::   :::     :::::::|||
 988 uLeuGlnAlaLysTyrProGluLeuPheAlaArgLysGlnIle....... 1002

1139 ACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCG 1188
                                         |||:::  ||||||
1003 .................................SerThrGlyValPro 1007

1189 ......CCGTCA..............................AACGG 1199
           ||||||                              |||||
1008 MetAspProSerValLysIleLysGlnTrpSerAlaLysTyrPheAsnGl 1024

1200 CAAAAATGTC......................................... 1209
      |  |||:::
1024 yThrAsnIleLeuGlyArgGlyAlaGlyTyrValLeuLysAspGlnAlaT 1041

1210 .............AAACTGGCAGACCAACGCCAC...........CCG 1233
                  :::::::::|||:::::::::           |||
1041 hrAsnThrTyrPheAsnIleSerAspAsnLysGluIleAsnPheLeuPro 1057

1234 AAGACA.......................GGCGTACCGTTTGACGGTAA 1259
      ||||||                      |||     :::||||||||
1058 LysThrLeuLeuAsnGlnAspSerGlnValGlyPheSerTyrAspGlyLy 1074

1260 AGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATA 1309
      ||||:::                      |||  |||:::|||
1074 sGlyTyr.................ValTyrTyrSerThr.......... 1081

1310 TTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGAT 1359
                 ||||||          :::|||||||  :::|||
1082 ...........SerGly.........TyrGlnAlaLysAsnThrPheIle 1091

1360 GCGAAA...CCGAGATGG...GAGGTTGATAGGAAG.............. 1389
      ::::::    :::|||       |||::::::
1092 SerGluGlyAspLysTrpTyrTyrPheAspAsnAsnGlyTyrMetValTh 1108

1390 .......................................CTTAATAAAT 1399
                                             |||   |
1108 rGlyAlaGlnSerIleAsnGlyValAsnTyrTyrPheLeuSerAsnGlyL 1125

1400 TGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGG 1449
      ||    |||:::   :::   ||||||   :::   |||
1125 euGlnLeuArgAspAlaIleLeuLysAsnGluAspGlyThr......... 1138

1450 AGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAAAT.. 1497
                  :::   ::::::      |||:::::::||||||
1139 ........TyrAlaTyrTyrGlyAsnAspGlyArgArgTyrGluAsnGl 1152

1498 ............AAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTG 1534
                   :::|||:::   ::::::||||||   :::|||:
1152 yTyrTyrGlnPheMetSerGlyVal...TrpArgHisPheAsnAsnGlyG 1168

1535 ATATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGT 1584
      ::::::::          ||||||    |||
1168 luMetSerValGlyLeuThrValIleAspGly.................. 1178

1585 GATGTACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTTTA 1634
                          |||::::::    |||:::   |||  ||
1179 ..................GlnValGlnTyrPheAspGluMetGly...Ty 1188

1635 TCAAGCGACAGTGGAA...ATTAAAAAGCCTGATGGAAGTTGGGAGGTGA 1681
      ||||||      :::    :::          |||||||:::   :::
1188 rGlnAlaLysGlyLysPheValThrThrAlaAspGlyLysIleArgTyrP 1205

1682 AAACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAA 1731
            |||::::::|||::::::
1205 heAspLysGlnSerGlyAsnMet........................... 1212

1732 GATTGGGATGAGGCTAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAG 1781
                                               ::::::::
1213 ........................................TyrArgAs 1215

1782 TAGA...ATAATGCTTAAGGATAATAAATGGCAGGGTACAAGTAAATCGG 1828
      :|||    |||      :::::::::||||||        :::::::::|
1215 nArgPheIleGluAsnGluGluGlyLysTrpLeuTyrLeuGlyGluAspG 1232

1829 GTATTAAAATAGAAGGATTTACCGAACCTAATAGAACAGCATATCCCATT 1878
      ||     :::    |||              |||    ||||||:::   ||
1232 lyAlaAlaValThrGlySerGlnThrIleAsn.GlyGlnHisLeuTyrPh 1248

1879 T 1879
     |
1248 e 1248
```

seq_name: SwissProt_40:VGP3_EBVA8

seq_documentation_block:

```
ID   VGP3_EBVA8     STANDARD;      PRT;   886 AA.
AC   Q07284;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Envelope glycoprotein GP340 (Membrane antigen) (MA).
GN   BLLF1.
OS   Epstein-barr virus (strain AG876) (Human herpesvirus 4), and
OS   Epstein-barr virus (strain P3HR-1) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=82830, 82829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AG876, AND P3HR-1;
RX   MEDLINE=93331716; PubMed=8393237;
RA   Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,
RA   Arrand J.R.;
RT   "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly
RT   conserved between virus types A and B.";
RL   Virology 195:578-586(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=P3HR-1;
RA   Klein K., Mueller-Lantzsch N.;
RL   Submitted (OCT-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L07922; AAA02783.1; -.
DR   EMBL; L07923; AAA02787.1; -.
DR   EMBL; X67776; CAA47986.1; -.
KW   Membrane; Glycoprotein; Antigen; Late protein.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   886 AA;  92388 MW;  4394F6130DECCA8A CRC64;

alignment_scores:
           Quality:  129.00              Length:    411
             Ratio:  0.694                 Gaps:     19
 Percent Similarity:  45.255    Percent Identity: 23.844

alignment_block:
 US-09-303-518D-463 x VGP3_EBVA8   ..

 Align seg 1/1  to: VGP3_EBVA8  from: 1  to: 886

      95 TTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATAC 144
         :::||||||  |||:::            ||||||  |||   :::
     379 IleSerGlyAlaPheAla............SerAsnArgThrPheAspIl 391

     145 CACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATA.... 190
          |||   |||:::     |||:::  |||:::::     |||
     391 eThrValSerGlyLeu...GlyThrAlaProLysThrLeuIleIleThrA 407

     191 ..............TCGGATTGGGAAACATACAAAGCCATCAGTTGGGCC 226
                       :::         |||:::|||:::|||
     407 rgThrAlaThrAsnAlaThrThrThrThrHisLysValIle......... 420

     227 ACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTC 276
                  |||:::::  |||      |||   ::::::|||
     421 .........PheSerLysAlaPro......GluSerThrThrThrSerPr 432

     277 CGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGC 326
          ::::::     ||||||      :::   :::    :::|||   :::|
     432 oThrLeuAsnThrThrGlyPheAlaAlaProAsnThrThrThrGlyLeuP 449

     327 CT.........CACATT..................CCGATTCTGACGAAG 349
          ||         |||:::                  |||    |||
     449 roSerSerThrHisValProThrAsnLeuThrAlaProAlaSerThrGly 465

     350 CCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGA 399
         |||:::|||        :::|||   :::       :::     ||||||
     466 ProThrValSerThrAlaAspValThrSerProThrProAlaGlyThrTh 482

     400 TACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTA 449
          :::::::::   |||   |||
     482 rSerGlyAlaSerProValThrProSer...................... 491

     450 TCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCG 499
            |||   |||:::   :::|||              |||    ||||||
     492 ..ProSerProArgAspAsnGly...........ThrGluSerLysAla 503

     500 TTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAA 549
                     |||   ::::::     ||||||   :::|||:::
     504 ProAspMetThrSerProThrSerAlaValThrThrProThrProAsnAl 520

     550 CGGCTTGCCGACCGTT.......................TCCACAATGC 575
                ||||||                           :::|||   |
     520 aThrSerProThrProAlaValThrThrProThrProAsnAlaThrSerP 537

     576 CGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCA...AACGCGCCA 622
          ||:::|||          |||          |||:::::   :::::|||

     537 roThrLeu......GlyLysThrSerProThrSerAlaValThrThrPro 551

     623 CCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC 672
               :::|||   |||   |||        :::       |||   |||::
     552 ThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAs 568

     673 AACGGCACTGCAGATATCGTCAAAAACATCA................... 703
         :::::::::          :::|||||||||
     568 nAlaThrIleProThrLeuGlyLysThrSerProThrSerAlaValThrT 585

     704 ...................TCGGCGCGGCAGGAGAAATTGTCGGCGCAG 733
                            |||              :::   |||   |||
     585 hrProThrProAsnAlaThrSerProThrValGlyGluThrSerProGln 601

     734 GCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTG......CTGTC 777
         |||          :::                    :::||||||   ||
     602 AlaAsnThrThrAsn...............HisThrLeuGlyGlyThrSe 613

     778 ATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAA 827
          |   |||      |||      |||||||||:::         |||   |
     613 rSerThrProValValThrSerProProLysAsnAlaThrSerAlaValT 630

     828 CGATTTGGCAG.........ATATGGCGCAACTCAAAGACTATGCCGCAG 868
          ||      |||                   ::::::::::::|||
     630 hrThrGlyGlnHisAsnIleThrSerSerSerThrSerSerMetSerLeu 646

     869 CAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATA 918
         :::||||||:::|||:::::::::   :::|||
     647 ArgProSerSerIleSerGluThrLeuSerProSerThrSerAspAsnSe 663

     919 GAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGG 968
                              |||   :::      |||:::   |||   :
     663 rThrSerHisMetProLeuLeuThrSerAlaHisProThrGlyGlyGluA 680

     969 AGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCA 1018
         ::::::::      :::   |||   ::::::::::::|||:::   :::
     680 snIleThrGlnValThrProAlaSerThrSerThrHisHisValSerThr 696

    1019 AGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTC 1068
         |||                   |||   :::   :::   :::      ||
     697 Ser..................SerProAlaProArgProGlyThrThrSe 707

    1069 AGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCA 1118
         |   :::         |||   :::   :::   :::|||:::|||
     707 rGlnAlaSerGlyProGlyAsnSerSerThrSerThrLysPro....... 721

    1119 TTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACA 1168
               |||::::::|||   |||   ::::::            :::   |||
     722 ..GlyGluValAsnVal...ThrLysGlyThrProProLysAsnAlaThr 736

    1169 TCACCTCCTCAACCGTGCCGCCGTCAAACGGCA 1201
         ||||||      |||      :::::::||||||
     737 SerProGlnAlaProSerGlyGlnLysThrAla 747
```

seq_name: SwissProt_40:VE2_HPV08

```
seq_documentation_block:
ID   VE2_HPV08     STANDARD;      PRT;   498 AA.
AC   P06422;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Regulatory protein E2.
GN   E2.
OS   Human papillomavirus type 8.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10579;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86200410; PubMed=3009874;
RA   Fuchs P.G., Iftner T., Weninger J., Pfister H.;
RT   "Epidermodysplasia verruciformis-associated human papillomavirus 8:
```

```
RT   genomic sequence and comparative analysis.";
RL   J. Virol. 58:626-634(1986).
CC   -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC       IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
CC       IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC       ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC       WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC       BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC       INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC       REPLICATION.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M12737; -; NOT_ANNOTATED_CDS.
DR   PIR; A03666; W2WL8.
DR   HSSP; P03122; 2BOP.
DR   InterPro; IPR001866; E2_N.
DR   InterPro; IPR000427; Early2_C.
DR   Pfam; PF00511; E2_C; 1.
DR   Pfam; PF00508; E2_N; 1.
DR   ProDom; PD000672; Early2_C; 1.
DR   ProDom; PD000678; E2_N; 1.
KW   Early protein; Transcription regulation; Activator; DNA-binding;
KW   Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ   SEQUENCE   498 AA;  56329 MW;  CC8003CE4DA69C48 CRC64;

alignment_scores:
          Quality:  128.50              Length:    312
            Ratio:  0.851                 Gaps:     16
 Percent Similarity:  48.397   Percent Identity: 27.885

alignment_block:
 US-09-303-518D-463 x VE2_HPV08   ..

 Align seg 1/1  to: VE2_HPV08  from: 1  to: 498

    401 ACGAACACCATCCCGCCGACGGCTATGACGGGCCAC...........AG 438
          |||:::::   |||||   :::        |||:::
    199 ThrSerSerThrProProGlySerProProGlyGlnAlaAspThrAspTh 215

    439 GGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGATATATACAGCTACGA 488
         ||||||      |||     :::|||    :::      |||:::  |||
    215 rAlaAlaLysThrProThrThrSerAlaAspSer......ThrSer.Arg 229

    489 CATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGAC........... 527
          ::::::|||         |||   |||    |||    :::
    230 GlnGlnArgSerProAlaLysGlnProGlnGlnThrGluThrLysGlyAr 246

    528 .........CGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTC 567
                  |||:::|||:::    ||||||        ||||||
    246 gArgTyrGlyArgArgProSerSerArgThr.........ArgProGlnL 260

    568 CACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACG 617
           :::   ||||||         :::||||||||||:::|||    ::::::
    260 ysGluGlnArgArg.........SerArgSerArgHisArgThrArgSer 273

    618 CGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC........ 659
         |||               |||:::          :::|||:::
    274 ArgSerArgSerLeuSerArgValArgAlaValGlySerThrThrValSe 290

    660 .........CGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAAC 699
                           ||||||    :::        |||    |||::::::
    290 rArgSerArgSerSerSerLeuThrLysAlaValArgProArgSerArgS 307

    700 ATCATCGGCGCGGCAGGAGAAATTGT...CGGCGCAGGCGATGCCGTGCA 746
           :::   |||||||||      ::::::         |||||||||:::
    307 erArgSerArgGlyArgAlaThrAlaThrSerArgArgArgAlaGlyArg 323

    747 GGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGC 796
         |||        ||||||     :::        ::::::           :::::
    324 GlySerProArgArgArgArgSerThrSerArgSerProSerThrAsnTh 340

    797 TTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCG 846
         :|||    |||:::|||    |||:::     :::|||               |||:
    340 rPheLysArgSerGlnArgGlyGlyGlyArgArgGlyArgGlyArgGlyS 357

    847 CAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAA 896
         ::       |||       |||||||||                 ||||||
    357 erArgGlyArgArgGluArgSerSerSerThr.........SerProThr 370

    897 CCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAG 946
         |||        |||:::|||     :::|||     :::        |||    |||
    371 ProThrLysArgSerArgGlyGluSerSerArg......LeuArgGlyVa 385

    947 CCATCCCCATCAA...............AGGGATTGGAGCTGTCCGGGGA 981
             |||    :::                     ::::::    |||:::
    385 lSerProSerGluValGlyArgSerValGlnSerValSerAlaLysHisT 402

    982 AAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGAT 1031
             |||||||||||                      :::     :::     |||
    402 hrGlyArgLeuGlyArg...........LeuLeuAspGluAlaIleAsp 414

   1032 GGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTG 1081
              :::     |||     :::|||::::::     :::              |||
    415 .ProProValIleLeuValArgGlyGluAlaAsnThrLeuLysCysPheA 431

   1082 CCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATC 1131
           :::     |||     ::::::|||  |||     |||              :::
    431 rgAsnArgAlaArgValArgTyr.ArgGlyLeuPhe......LysTyrPh 445

   1132 CGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAAC 1181
              |||||||||    |||::::::    :::|||:::
    445 eSerThrThrTrpSerTrpValAlaGlyAspSerThrGlu.......... 458

   1182 CGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACC 1231
             |||              ::::::|||              |||:::|||
    459 .....ArgLeuGlyArgSerArgMetLeuIleLeuPheThrSerAlaGly 473

   1232 CGAAGACAGGCGTACCGTTTGACGGTAAAG 1261
         :::|||:::    :::      |||||||||
    474 GlnArgLysAspPheAspGluThrValLys 483

seq_name: SwissProt_40:DRPL_RAT

seq_documentation_block:
ID   DRPL_RAT       STANDARD;      PRT;  1183 AA.
AC   P54258;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN   DRPLA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum, and Striatum;
RX   MEDLINE=97317138; PubMed=9173996;
RA   Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA   Ashworth R.G., Ross C.A.;
RT   "Cloning and expression of the rat atrophin-I (DRPLA disease gene)
RT   homologue.";
RL   Neurobiol. Dis. 2:129-138(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX   MEDLINE=96081227; PubMed=8541849;
RA   Schmitt I., Epplen J.T., Riess O.;
RT   "Predominant neuronal expression of the gene responsible for
RT   dentatorubral-pallidoluysian atrophy (DRPLA) in rat.";
RL   Hum. Mol. Genet. 4:1619-1624(1995).
CC   -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC       MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC   -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC       (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U31777; AAA80337.1; -.
DR   EMBL; X89453; CAA61623.1; -.
DR   InterPro; IPR002951; Atrophin.
DR   PRINTS; PR01222; ATROPHIN.
FT   DOMAIN      165    171       POLY-PRO.
FT   DOMAIN      303    306       POLY-PRO.
FT   DOMAIN      377    383       POLY-SER.
FT   DOMAIN      387    391       POLY-SER.
FT   DOMAIN      440    446       POLY-PRO.
FT   DOMAIN      477    480       POLY-HIS.
FT   DOMAIN      481    489       POLY-GLN.
FT   DOMAIN      502    505       POLY-PRO.
FT   DOMAIN      562    572       POLY-SER.
FT   DOMAIN      702    705       POLY-PRO.
FT   CONFLICT    455    455       N -> S (IN REF. 2).
FT   CONFLICT    594    594       F -> L (IN REF. 2).
FT   CONFLICT    689    689       P -> R (IN REF. 2).
FT   CONFLICT    717    717       T -> M (IN REF. 2).
FT   CONFLICT    737    737       A -> V (IN REF. 2).
FT   CONFLICT    965    965       MISSING (IN REF. 2).
SQ   SEQUENCE   1183 AA;  124778 MW;  7FB9928DCADF9B1F CRC64;

alignment_scores:
           Quality:  126.50              Length:    509
             Ratio:  0.591                 Gaps:     31
 Percent Similarity:  42.043    Percent Identity: 24.165

alignment_block:
 US-09-303-518D-463 x DRPL_RAT   ..

 Align seg 1/1  to: DRPL_RAT  from: 1  to: 1183

      12 CCGCAAAATATCCCTTATTCTGTCCATACTGGC..............AG 46
         |||::::::||||||   :::   ::::::|||               ||
     171 ProAspSerIleProArgGlnProGluSerGlyPheGluProHisProSe 187

      47 TGTGCCTGCCGATGCA.......................TGCACACGCC 72
         |||||||        :::                        :::||||
     187 rValProProThrGlyTyrHisAlaProMetGluProProThrSerArgL 204

      73 TCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCG...... 116
         ||      |||                  |||   |||:::   |||
     204 euPheGlnGlyPro...........ProProGlyAlaProProProHis 216

     117 ..................TCAGCATTTCGAACCCGACGGGAAATACCACC 148
                           ||||||                :::   |||||
     217 ProGlnLeuTyrProGlySerAlaGlyGlyGlyValLeuSerGlyProPr 233

     149 TATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAA...CGGCCATATCGGA 195
         |:::        :::|||||||||:::::::::::          |||
     233 oMetGlyProLysGlyGlyAlaAlaAlaSerSerValGlyProProSerG 250

     196 TTGGGAAA...CATACAAAGCCATCAGTTGGGCCACCTGATGATTCAACA 242
            ||||||    |||        |||::::::        |||    :::|||:::
     250 lyGlyLysGlnHisProProProThrThrProIleProIleSerSerSer 266

     243 GGCGG............CCGTTGAAGGAAATATCGGCTACATTGTCCGC 279
         |||                   |||   |||        :::::
     267 GlyAlaSerGlyAlaProProAlaLysProProAsnThr........... 279

     280 TTTTCCGATCACGGGCACAAAT......TCCATTCGCCCTTCGACAACCA 323
               |||:::    :::   |||       |||      |||    :::|||:
     280 ....ProValGlyAlaGlyAsnLeuProSerAlaProProProAlaThrP 295

     324 TGCCTCACATT...CCGATTCTGACGAAGCCG......GTAGTCCCGTTG 364
         ::||||||:::    |||    |||      |||      :::    ||||||
     295 heProHisValThrProAsnLeuProProProProAlaLeuArgProLeu 311

     365 ACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCC 414
            ::::::|||   ::::::          |||        ::::::    |||||
     312 ...AsnAsnAlaSerAlaSerProProGlyMetGlyAlaGlnProIlePr 327

     415 G..............CCGACGGCTATGACGGGCCACAGGGCGGCGGCTA 449
         |                |||   ||||||   |||:::       :::::::
     327 oGlyHisLeuProSerProHisAlaMet...GlyGlnGlyMetSerGlyL 343

     450 TCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCG 499
         ::|||              :::|||            |||    :::
     343 euProProGlyProGluLysGlyProThrLeuAlaProSerProHisPro 359

     500 TTGCCCAAAATATCCGCCTCAACC..................TGACCGAC 531
         ||||||       |||:::|||:::                       |||
     360 LeuProProAlaSerSerSerAlaProGlyProProMetArgTyrProTy 376

     532 AACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGC 581
           :::::::::   ::::::::::::    :::    |||:::       |||
     376 rSerSerCysSerSerSerSerValAlaAlaSerSerSerSerSerAla. 392

     582 TATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACA 631
                                 ||||||::::::              :::|||
     393 ....................AlaThrSerGlnTyrProAlaSerGlnThr 402

     632 GCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACT 681
            ||||||:::                 ||||||          |||:::::
     403 LeuProSerTyrProHisSerPheProProPro.........ThrSerMe 416

     682 GCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGC 731
         ::::::|||::::::          :::       |||    :::|||     :::|
     416 tSerValSerAsnGlnProProLysTyrThrGlnProSerLeuProSerG 433

     732 AGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGC 781
         ||||||:::
     433 lnAlaVal........................................ 435

     782 ACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGAT 831
               |||         ||||||
     436 ......TrpSerGlnGlyProProProPro.........ProProProTy 447

     832 TTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAG...CCATCCG 878
            ::::::      |||::::::|||    |||           |||    :
     447 rGlyArgLeuLeuProAsnAsnAsnThrHisProGlyProPheProProT 464

     879 CGATTGGGCAGTCCAAAACC......CCAATGCCGCACAAGGCATAGAAG 922
         ::    |||||||||      :::       |||    |||  ::::::|||:::
     464 hrGlyGlyGlnSerThrAlaHisProProAlaPro.AlaHisHisHisHi 480

     923 CCGTCAGCAATATCTT................................ 938
          :::||||||
     480 sGlnGlnGlnGlnGlnProGlnProGlnProGlnProGlnGlnHisHisH 497

     939 .......................TATGGCAGCCATCCCCATCAAAGGGAT 965
                                 |||::::::||||||       :::
     497 isGlyAsnSerGlyProProProProGlyAlaTyrProHisProLeuGlu 513

     966 TGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCT. 1014
            |||                        :::   ||||||  |||||||||
```

```
 514 SerSer....................AsnSerHisHis.AlaHisProT 523

1014 ............................................... 1014

 523 yrAsnMetSerProSerLeuGlySerLeuArgProTyrProProGlyPro 539

1015 ..........................GTCAAGCGGTCGCAGATGGGCGC 1037
                               |||:::  ||||||  |||
 540 AlaHisLeuProProSerHisGlyGlnValSerTyrSerGlnAlaGlyPr 556

1038 GATCGCATTGCCGAAAGGGAAATCCGCC...GTCAGCGACAATTTTGCCG 1084
        :::  |||  ::::::|||:::      |||  :::  ::::
 556 oAsnGlyProProValSerSerSerSerAsnSerSerGlySerSerSerG 573

1085 ATGCGGCATAC..................................... 1095
     ::|||||||||
 573 lnAlaAlaTyrSerCysSerHisProSerSerSerGlnGlyProGlnGly 589

1096 GCCAAATACCCGTCCCCTTACCATTCCCGAAATA......TCCGTTCAAA 1139
     |||:::||||||  |||   :::|||  |||     |||      |
 590 AlaSerTyrProPhePro.ProValProProIleThrThrSerSerAlaT 606

1140 CTTGGAGCAGCGTTACGGCA.........................AAG 1162
     ||  |||:::|||  |||                          |||
 606 hrLeuSerThrValIleAlaThrValAlaSerSerProAlaGlyTyrLys 622

1163 AAAACATCACCTCCTCAACCG 1183
        :::|||||||||  |||
 623 ThrAlaSerProProGlyPro 629
```

seq_name: SwissProt_40:AGA1_YEAST

```
seq_documentation_block:
ID   AGA1_YEAST     STANDARD;      PRT;   725 AA.
AC   P32323;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   A-agglutinin attachment subunit precursor.
GN   AGA1 OR YNR044W OR N3431.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91304412; PubMed=2072914;
RA   Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT   "The AGA1 product is involved in cell surface attachment of the
RT   Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RL   Mol. Cell. Biol. 11:4196-4206(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Pohl T.M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC       SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC       CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC       RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC       AGGREGATION DURING MATING.
CC   -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC       A CORE SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M60590; AAA34382.1; -.
DR   EMBL; Z71659; CAA96325.1; -.
DR   PIR; S17031; S17031.
DR   PIR; A41258; A41258.
DR   SGD; S0005327; AGA1.
KW   Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
KW   Pheromone response.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    725       A-AGGLUTININ ATTACHMENT SUBUNIT.
FT   DOMAIN       53    493       2 X APPROXIMATE REPEATS.
FT   REPEAT       53    149       1-1.
FT   REPEAT      395    493       1-2.
FT   DOMAIN      182    307       18 X APPROXIMATE TANDEM REPEATS, SER/THR-
FT                                RICH.
FT   REPEAT      182    188       2-1.
FT   REPEAT      189    195       2-2.
FT   REPEAT      196    202       2-3.
FT   REPEAT      203    209       2-4.
FT   REPEAT      210    216       2-5.
FT   REPEAT      217    223       2-6.
FT   REPEAT      224    230       2-7.
FT   REPEAT      231    237       2-8.
FT   REPEAT      238    244       2-9.
FT   REPEAT      245    251       2-10.
FT   REPEAT      252    258       2-11.
FT   REPEAT      259    265       2-12.
FT   REPEAT      266    272       2-13.
FT   REPEAT      273    279       2-14.
FT   REPEAT      280    286       2-15.
FT   REPEAT      287    293       2-16.
FT   REPEAT      294    300       2-17.
FT   REPEAT      301    307       2-18.
SQ   SEQUENCE   725 AA;  73353 MW;  70420C853B0B01F8 CRC64;
```

```
alignment_scores:
           Quality:  125.00            Length:    431
             Ratio:  0.551               Gaps:     18
 Percent Similarity:  52.668   Percent Identity: 20.882

alignment_block:
 US-09-303-518D-463 x AGA1_YEAST   ..

 Align seg 1/1  to: AGA1_YEAST  from: 1  to: 725

      20 TATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATGCATGCAC.. 67
          |||      :::||||||           :::       :::   :::
      67 TyrThrThrTrpCysProLeuThrValSerThrSerSerAlaAlaGluIl 83

      68 .ACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCG 116
          :::|||:::|||      :::      ||||||   |||||||||||:
      83 eSerProSerIleSerTyrAlaThrThrLeuSer...ArgPheSerThrL 99

     117 TCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGG 166
         :::::::|||:::  :::  :::                 :::::::::
      99 euThrLeuSerThrGluValCysSerHisGluAlaCysProSerSerSer 115

     167 AGCTTGCCNAGCGCAACG......GCCATATCGGATTGGGAAACATAC.. 208
         :::||||||:::::|||     :::   |||   :::   :::|||
     116 ThrLeuProThrThrThrLeuSerValThrSerLysPheThrSerTyrIl 132

     209 ..................AAAGCCATCAGTTGGGCCACCTGATGATTCA 239
                             |||||||||
     132 eCysProThrCysHisThrThrAlaIleSer.................. 142

     240 ACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATC 289
                   |||:::|||:::::::::|||               :::
     143 ........SerLeuSerGluValGlyThrThr...........ThrVal 152

     290 ACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGAT 339
         :::::::::::::::|||:::||||||||::::::         |||::
     153 ValSerSerSerAlaIleGluProSerSerAlaSerIleIleSerProVa 169
```

```
 340 TCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCA 389
     :  |||:::   :::            |||:::|||      |||::::::
 169 1...ThrSerThrLeuSerSerThrThrSerSerAsnProThrThrThrS 185

 390 TTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGG 439
        ::::::|||:::|||   |||   ::::::   |||:::
 185 erLeuSerSerThrSerThrSerProSerSerThrSerThrSerProSer 201

 440 GCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGAC 489
     :::::::::                ::::::   :::    |||:::::::::
 202 SerThrSer.........ThrSerSerSerSerThrSerThrSerSerSe 215

 490 ATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAG 539
     :        :::       :::  |||:::||||||   |||::::::::::
 215 rSerThrSerThrSerSerSerSerThrSerThrSerProSerSerThrS 232

 540 CACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGA 589
     ::   :::::::::|||   ::::::|||:::        :::
 232 erThrSerSerSerLeuThrSerThrSerSerSerSerThrSerThrSer 248

 590 CGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAG 639
     :::::::::     :::|||:::|||::::::           :::|||:::|||||
 249 GlnSerSerThrSerThrSerSerSerSer...ThrSerThrSerProSe 264

 640 CTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATAT 689
     |    |||:::   :::                 :::   ||||||:::   :::   |
 264 rSerThrSerThrSer.........SerSerSerThrSerThrSerProS 278

 690 CGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATG 739
     ||||||||::::::||||||    :::    :::    |||:::    :::
 278 erSerLysSerThrSerAlaSerSerThrSerThrSerSerTyrSerThr 294

 740 CCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTG 789
                       :::   :::   :::|||   |||   :::
 295 ..............SerThrSerProSerLeuThrSerSerSerProTh 306

 790 GGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGA 839
      :::                                    |||||||||      :::
 306 rLeu..........................AlaSerThrSerProSerS 314

 840 TATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAG 889
                    |||:::|||:::   ::::::   |||:::::::|||:::
 314 erThrSerIleSerSerThrPheThrAspSerThrSerSerLeuGlySer 330

 890 TCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTT 939
      |||                                |||    :::    :::::::::|||||
 331 Ser..................IleAlaSerSerSerThrSerValSerLe 341

 940 ATGGCAGCCATCC...CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATA 986
     |::::::||||||    |||          ::::::          :::::::::|||:
 341 uTyrSerProSerThrProValTyrSerValProSerThrSerSerAsnV 358

 987 CGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCG 1036
     ::|||
 358 alAla.......................................... 359

1037 CGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT 1086
                                  ||||||   |||              ::
 360 .......................ThrProSerMetThrSerSerThrVa 368

1087 GCGGCATACGCCAAATACCCGT........................... 1108
     ::::   |||   :::::::::
 368 lGluThrThrValSerSerGlnSerSerSerGluTyrIleThrLysSerS 385

1109 .....CCCCTTACCATTCCCGAAATATCCGTTCAAACTTGG......... 1144
                  |||||||||::::::|||::::::|||:::
 385 erIleSerThrThrIleProSerPheSerMetSerThrTyrPheThrThr 401

1145 ...AGCAGCGTTACGGCAAAGAAAACATCA...CCTCCTCAACCGTGCCG 1188
        |||:::||||||:::      |||:::       |||            ::
 402 ValSerGlyValThrThrMetTyrThrThrTrpCysProTyrSerSerGl 418

1189 CCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACC 1231
        :   :::|||:::   :::::::::   :::   ::::::|||
 418 uSerGluThrSerThrLeuThrSerMetHisGluThrValThr 432
```

seq_name: SwissProt_40:HIG_DROME

```
seq_documentation_block:
ID   HIG_DROME      STANDARD;      PRT;   958 AA.
AC   Q09101;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Locomotion-related protein Hikaru genki precursor.
GN   HIG.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Head;
RX   MEDLINE=93213498; PubMed=8461133;
RA   Hoshino M., Matsuzaki F., Nabeshima Y.-I., Hama C.;
RT   "Hikaru genki, a CNS-specific gene identified by abnormal locomotion
RT   in Drosophila, encodes a novel type of protein.";
RL   Neuron 10:395-407(1993).
CC   -!- FUNCTION: HAS A ROLE IN THE DEVELOPMENT OF CNS FUNCTIONS INVOLVED
CC       IN LOCOMOTOR ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN PCC NEURONS AND NEUROBLASTS
CC       IN THE PROCEPHALIC NEUROGENIC REGION IN THE CENTRAL NERVOUS
CC       SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING AND/OR AFTER NEURONAL
CC       DIFFERENTIATION AND DURING CELL SPECIFICATION OR AXOGENESIS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D13884; BAA02984.1; -.
DR   EMBL; D13885; BAA02985.1; -.
DR   EMBL; D13886; BAA02986.1; -.
DR   EMBL; D13887; BAA02987.1; -.
DR   HSSP; P10998; 1VVD.
DR   FlyBase; FBgn0010114; hig.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00409; IG; 1.
KW   Glycoprotein; Alternative splicing; Immunoglobulin domain; Repeat;
KW   Sushi; Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    958       LOCOMOTION-RELATED PROTEIN HIKARU GENKI.
FT   DOMAIN      630    709       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      713    769       SUSHI 1.
FT   DOMAIN      772    828       SUSHI 2.
FT   DOMAIN      831    891       SUSHI 3.
FT   DOMAIN      893    952       SUSHI 4.
FT   SITE        318    320       CELL ATTACHMENT SITE.
FT   DISULFID    714    755       BY SIMILARITY.
```

```
FT   DISULFID    741    768       BY SIMILARITY.
FT   DISULFID    773    814       BY SIMILARITY.
FT   DISULFID    800    827       BY SIMILARITY.
FT   DISULFID    832    877       BY SIMILARITY.
FT   DISULFID    863    890       BY SIMILARITY.
FT   DISULFID    894    939       BY SIMILARITY.
FT   DISULFID    922    952       BY SIMILARITY.
FT   CARBOHYD    376    376       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    525    525       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    605    605       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    752    752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    789    789       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    529    553       MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT   VARSPLIC    892    958       MISSING (IN ISOFORM 1 AND ISOFORM 2).
SQ   SEQUENCE   958 AA;  107027 MW;  4161258E85ABC764 CRC64;

alignment_scores:
            Quality:  124.50          Length:    571
              Ratio:  0.502             Gaps:     27
 Percent Similarity:  43.433   Percent Identity: 20.841

alignment_block:
 US-09-303-518D-463 x HIG_DROME   ..

 Align seg 1/1  to: HIG_DROME  from: 1  to: 958

    487 GACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCG 536
         |||     |||      ::::::  |||||||||         |||||
    139 AspHisHisGlyAspProGluAspLysArgLeuAsnAla......AsnAr 153

    537 CAGCACCGGACAACGG.................CTTGCCGACCGTTTCC 568
         |:::   |||::::|||                     |||::::
    153 gLysGlnGlyLysArgArgArgAlaGlySerGlyArgArgArgArgIleG 170

    569 ACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGC 618
         ::|||   :::::::::      :::|||   |||:::    :::||||||
    170 luAsnGluAsnGlyGlnThrGlyArgGlyArgGlySerArgTyrLysArg 186

    619 ..............GCCACCCGATACAGCCCCGAGCTGGACAGATCGGG 653
                               |||:::   |||||||||  |||||| . ::
    187 HisAlaIleLeuHisAspThrGluAlaSerProGluThrAspArgTrpAl 203

    654 CAATGCCGCCGAAGCCTTCAACGGC.................ACTGCAG 685
         :::::::      |||    :::|||                    ||||
    203 aGlySerLysLeuAlaAlaGluGlyAspValTyrTyrValHisIleAlaA 220

    686 ATATCGTCAAAAACATCATCGGCGCGGCAGGAG................ 718
         ||||||::: |||:::    :::    :::::::|||
    220 spIleLeu.LysSerArgGluProAsnArgGluLeuLysSerLysLeuHi 236

    719 .......AAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGG 761
                 |||:::::::|||:::           |||     :::    :::
    236 sLysLeuLysMetLysAlaArgLeuAsnLysCysLeuAlaGluGlyGlyL 253

    762 CTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACA 811
            :::          :::||||||       :::        ||||||
    253 ysGlu.........LysCysThrArgLeuLeuLysLysLysProLys... 265

    812 AGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTAT 861
                                                 :::|||::::
    266 ....................................LysLysValVa 269

    862 GCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACA 911
         :   ::::::    :::          :::::::|||    |||        ::::
    269 lGluLysGluGlnThrLeuLysLysGluLysLysPheProLysGluGluG 286

    912 AGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAG 961
         :::::    |||    ::::::::::::          |||         :::
    286 lnSer...LysGluLysValLeuGluAsnGlyGlnThrProLysGluAsp 301

    962 GGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCAC....... 1004
                 ||||||||:::          :::||||||   |||  ||||||
    302 GluLeuGluLeuAspHisProGluThrAlaAlaAla.HisHisArgArgA 318

   1005 ..GGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAA 1052
          |||   |||       |||        |||    |||||||||        :::
    318 rgGlyAspSerHisAlaAlaGluLeuAspGlnArgAspArgSerProArg 334

   1053 AGGGAAATCCGCCGTCAGCGACAATTT...................... 1079
           :::    |||||||:::     :::|||
    335 TrpArgArgArgArgSerThrGluPheLysGlyAspLeuGlyGlnLeuPr 351

   1080 .TGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAAT 1128
           :::    ||||||              |||    ||||||      :::::::|||
    351 oProGluSerGlyIleGlyProGluProGluProLeuAla.AspGlnAsn 367

   1129 ATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTC 1178
         :::::::::               |||:::|||||||:::::::::    ::::::
    368 LeuLysAsp......LeuGlnGlnTyrGlyAsnGlnSerSerSerAlaAr 382

   1179 AACCGTG.......................................... 1185
          ::::::
    382 gValAlaLeuLeuTrpGlnArgValLysArgLysSerGlyArgThrAlaG 399

   1186 ..............CCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGAC 1221
                          |||    ::::::|||     :::     :::     :::
    399 lyAlaLeuSerArgProLysGlyGlyGlyAspSerSerSerLysThrThr 415

   1222 CAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTT..... 1266
         :::|||       |||     |||:::            |||:::
    416 SerArgLysAspLys...GlyIleTyrAspGluGluAlaGlyTyrThrPr 431

   1267 ...............CCGAATTTTGAGAAGCACGTGAAATATGATACGA 1300
                           |||:::|||:::::::::     :::     |||     :
    431 oIleHisProAspAspProGluPheAspGluGluGluGluGluAspGluG 448

   1301 AGCTCGATATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCT 1350
         ::::::||||||                      :::     :::
    448 luValAspIle....................LeuGlnGlnPheThrGlu 457

   1351 GTGTTTGATGCGAAACCGAGATGGGAGGTT...........GATAGGAA 1388
         |||     :::     :::           |||:::               ||||||::
    458 ValSerGluIleArgPheProGlyGluIleGlyProMetGlyAspArgAr 474

   1389 GCTTAATAAATTG.................................... 1401
         :|||    |||:::
    474 gLeuCysLysIleArgCysValLysGlyLysTrpValGlyProLeuCysA 491

   1402 ..ACAACTCGTGAGCAGGTGGAGAAAAATGTT......CAGGAAACGAGA 1443
           |||::::::|||:::     :::     ||||||        |||
    491 laThrAsnGluGluAspAspAsnGlyAsnValLysPheGlnProLeuTyr 507

   1444 AGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGA 1493
         :::        :::     :::::::::     :::|||          :::::::
    508 LysSerCysHisValAsnArgIleProSerHisLeuLeuLeuSerTyrAr 524

   1494 AAAT.....................................AAAACAG 1504
         :|||                                       ||||||
    524 gAsnIleSerValThrProIleProProAsnArgGlyTrpArgLysThrA 541

   1505 GGTTAGATTTTAATCATTTTATAGGTGGT...GATATCAATAAGAAAGGC 1551
          |||:::     :::     ::::::::::::     :::||||||
    541 rgLeuSerLysSerThrLeuLeuSerAsnThrGluIleAsnValGlyTrp 557

   1552 ACAGTAACAGGAGGGCATAGTCTAACCCGT................... 1581
           :::          ||||||||||||
    558 AspLeuProHisGlyHisSerLeuGlnAlaArgCysGlnGluLeuGlyIl 574

   1582 ............GGTGATGTACGGGTGATA.................. 1599
                         |||:::    ||||||:::
    574 eTyrLysLeuLeuGlyGluSerArgValLeuCysSerAsnGlyLeuTrpA 591

   1600 ............................................CAA 1602
                                                      :::
```

```
   591 laProArgMetProSerCysValProThrThrValLeuThrAsnTyrSer 607

  1603 CAAACCTCGGCACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAAAT 1652
       :::   |||||||||                        :::::::::||
   608 GluAspSerAlaPro.......................SerIleArgIl 616

  1653 TAAAAAGCCTGATGGAAGTTGGGAGGTGAAAACGAAAAAAGGTGGGAAAG 1702
       ||||        :::||||||  :::   :::      :::|||     :
   616 eLysIlePheAsnGlySerHisSerPheGlu...ProSerGlyValMetA 632

  1703 TGATGACCAAGCACACCATGTTCCCAAAAGATTGGGATGAGGCTAGAATT 1752
       :::::       |||:::              |||          |||:::
   632 laValProProHisSerThrValLeuMetAspCysMetTyrProArgVal 648

  1753 AGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGATAA 1802
       |||:::              :::|||   |||     |||      :::::
   649 ArgGlyThrProGluTrpSerTrpThrSerTrpTyrMetGlnTyrSerTh 665

  1803 TAAATGG 1809
       :   |||
   665 rGlyTrp 667
```

seq_name: SwissProt_40:YQU3_CAEEL

seq_documentation_block:

```
ID   YQU3_CAEEL     STANDARD;      PRT;  1251 AA.
AC   Q09550;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical 133.5 kDa protein F26C11.3 in chromosome II.
GN   F26C11.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Matthews P.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z47072; CAA87369.1; -.
DR   WormPep; F26C11.3; CE01561.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 1.
DR   SMART; SM00032; CCP; 1.
KW   Hypothetical protein.
FT   DOMAIN      120    414       SER/THR-RICH.
SQ   SEQUENCE   1251 AA;  133498 MW;  2B959ECA03B9954A CRC64;
```

alignment_scores:

```
           Quality:  124.50              Length:    413
             Ratio:  0.635                 Gaps:     18
Percent Similarity:  47.458    Percent Identity: 20.339
```

alignment_block:

```
US-09-303-518D-463 x YQU3_CAEEL   ..

Align seg 1/1  to: YQU3_CAEEL  from: 1  to: 1251

   140 AATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCAT 189
        |||    :::    :::::::::::::::::::::   |||:::|||:::::
   208 AsnLeuSerThrThrThrSerSerSerThrMetLeuSerSerThrThrLe 224

   190 ATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCA 239
       :            ||||||
   224 uLeuThrThrGluThr.................................. 229

   240 ACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATC 289
                     :::|||   |||:::|||   |||:::
   230 ........GluThrArgGluSerSerSerThrGlySerThr......Gln 241

   290 ACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGAT 339
       |||   |||    |||    :::||||||||||||:::
   242 ThrThrThrProSerThrGluProSerThrThrIle.............. 253

   340 TCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCA 389
                     ::::::|||:::   :::|||:::   ::::::|||:
   254 ............ThrThrProMetGluGlnSerSerThrValSerSerV 266

   390 TT.........................GGGACGGATACGAACACCATC 412
       ::                            :::::::::::::|||:::
   266 alGlnLysThrArgThrSerGluAspLysProSerSerSerThrThrVal 282

   413 CCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAA 462
       |||    :::|||   |||:::::   :::::::::   |||:::    ::
   283 ProThrSerAlaSerThrSerGluSerSerThrSerSerProMetAlaGl 299

   463 GGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATAT 512
       ::::   :::       ::::::||||||   ::::::    |||      |
   299 uThrSerSer......SerSerThrThrSerGlnSerSerPro...AlaS 313

   513 CCGCCTCAACCTGACCGACAACC............GCAGCACCGGACAAC 550
       |||:::||||||    |||    :::            :::::::|||   :::
   313 erThrSerThrValProGluSerSerThrValGlySerThrProThrThr 329

   551 GGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTA 600
       ||||||    |||:::||||||                       :::|||
   330 GlyLeuThrThrLeuSerThr......................AsnGluGl 339

   601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATC 650
        :::|||:::::::::::::           ::::::   ::::::
   339 nSerThrSerThrSerSerGlyGlyHisSerThrSerThrPheGlyThrT 356

   651 GGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACA 700
        :::          |||:::   ||||||    :::      :::|||    :::
   356 hrSerGluThrProGluThrSerThrAspPheThrAlaThrSerThrSer 372

   701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
       ||||||:::    ::::::      |||::::::|||
   373 SerSerSerAspSerSerThrGlnSerSerAsnAla.............. 384

   751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTC 800

   384 .................................................. 384

   801 CACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 850
                        ::::::|||||||||            :::|||
   385 .................GlnThrSerThrIle............GluAsn 391

   851 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900
       ::::::|||       :::          ||||||    :::    |||:::|||||
   392 GlySerThrThrThrAsnPheThrSerAlaProSerThrSerSerThrPr 408

   901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCAT 950
       |                     |||::::::            |||          :
   408 oAla..............ThrProThrThrThrTyrAsnTrpProThrG 420

   951 CCCCATCAAAGGGATTGGAGCTGTCCGGGGAA..............AAT 985
       ::    :::      :::   |||||||||                  :::
   420 lyGlyThrThrTrpMetLeuProSerGlyGluIleValGlnAlaHisLys 436

   986 ACGGCTTGGGCGGCATCACGGCACATC...CTGTCAAGCGGTCGCAGATG 1032
       |||    :::::::::   :::::::::    :::    |||
   437 ThrPhePheSerSerPheAsnTyrPheGlnValLeuSerGluSerLeuIl 453
```

```
1033 GGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTGC 1082
       |||      :::|||                 ::::::::::::::
 453 eAlaTyrProAsnCys.................ThrThrValLeuMetG 464

1083 CGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCC 1132
       :::    ::::::|||:::|||:::
 464 lnLeuIleTyrAsnProArgThrLys....................... 472

1133 GTTCAAACTTGGAGCAGCGTTACGGCA..................AAGAA 1164
        :::|||    :::::::::|||:::                   |||
 473 ...GluThrArgThrGluIleThrSerAspAlaGluGlyCysLysLysTh 488

1165 AACATCACCTCCTCAACCG............TGCCGCCGTCAAACGGCAA 1202
       :::|||    |||   |||               :::      ||||||
 488 rSerSerThrProThrProSerSerThrSerValHisSerThrThrAlaT 505

1203 AAATGTCA.......................AACTGGCAGACCAACGCC 1228
           |||                      ||||||   |||::::::
 505 hrProSerThrThrProGlyThrThrThrTyrAsnTrpProThrGlyGly 521

1229 ACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTC 1267
      |||   |||:::                 :::   ||||||
 522 ThrThrArgMetLeuProSerGlyGluIleValGlyPhe 534
```

seq_name: SwissProt_40:V70K_TYMVC

```
seq_documentation_block:
ID   V70K_TYMVC     STANDARD;      PRT;   628 AA.
AC   P28478;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   69 kDa protein.
OS   Turnip yellow mosaic virus (isolate TYMC).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX   NCBI_TaxID=31751;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92119261; PubMed=1731998;
RA   Dreher T.W., Bransom K.L.;
RT   "Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a
RT   cDNA-based clone with verified infectivity.";
RL   Plant Mol. Biol. 18:403-406(1992).
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X16378; CAA34414.1; -.
DR   PIR; S19150; S19150.
SQ   SEQUENCE   628 AA;  69057 MW;  0DF0C64E8ECBC6DC CRC64;

alignment_scores:
           Quality:  124.00           Length:     446
             Ratio:  0.642              Gaps:      28
 Percent Similarity:  43.274   Percent Identity: 25.336

alignment_block:
 US-09-303-518D-463 x V70K_TYMVC  ..

 Align seg 1/1  to: V70K_TYMVC  from: 1  to: 628

   8 TTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCG 57
      |||        :::||||||      |||        :::         ||
  78 PheSerProLeuHisProLeuValPheProGlyHisHisSerGlnLeuAr 94

  58 ATGCATGCAC...ACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCA 104
      |   :::|||    |||   |||:::    ||||||    |||
  94 gHisValHisGluThrGlnGlnValGlnGlnThrCysPro.......... 107

 105 GGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATT.. 152
           :::    :::::::::|||:::     :::|||:::||||||
 108 .....GlyLysLeuLysLeuSerGly.AlaGluGluLeuProProAlaPr 122

 153 ....CGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTG 198
         ||||||::::::                      ||||||    |||:::
 122 oGlnArgGlnHisSerLeuProLeuHisIleThrArgProSerArgPheP 139

 199 GGAAACAT.........ACAAAGCCATCAGTTGGGCCA......CCTGAT 233
         |||             :::|||:::|||    |||         ||||||
 139 roHisHisPheHisAlaArgArgProAspValLeuProSerValProAsp 155

 234 GATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTT 282
                                              |||    |||:::
 156 ..................................HisGlyProValLe 160

 283 .............CCGATCACGGGCACAAATTCCATTCGCCCTTCGACAA 320
                  |||   |||:::::::::::     |||    :::|||
 160 uAlaGluThrLysProArgThrSerValArgGlnProArgSerAlaThrA 177

 321 CCATGCCTCACATT...CCGATTCTGACGAAGCCGGTAGTCCCGTTGACG 367
        |||    :::    |||||||||             ||||||    :::
 177 rgGlyProSerPheArgProIleLeuLeuProLysValValHisValHis 193

 368 GATTCAGCCTTTACCGCATCCATT.........GGGACGGATACGAACAC 408
     |||:::           :::|||:::          |||:::    ::::::
 194 AspAspProProHisSerSerLeuArgProArgGlySerArgSerArgGl 210

 409 CATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCC 458
       :::    ||||||:::
 210 nLeuGlnProThrValArgArg............................ 217

 459 CAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAA 508
                                                   |||
 218 ..........................................ProLeu 219

 509 ATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCC 558
      ::::::    :::             :::            |||         ||
 220 LeuAlaProAsnGlnPheHisSerProArgGlnPro............Pr 232

 559 GACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGG 608
      |    :::|||        |||::::::
 232 oProLeuSerAspAspProGlyIleLeu...................... 241

 609 ATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATG 658
              :::|||         ||||||:::    |||     |||
 242 ........GlyProArgProLeuAlaProAsnSerThr...ArgAspPro 254

 659 CCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGC 708
     ||||||:::|||    |||                   |||:::|||    :::
 255 ProProArgProIleThr......ProGlyProSerAsnThrHisGlyLe 269

 709 GCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGA 758
       |||          ::::::       :::|||   |||    |||
 269 uArgProLeuSerValLeuProArgAlaSerProArgArg.......... 282

 759 AGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAA 808
                                                :::|||
 283 ..............................GlyLeuLeuProAsnPro 288

 809 ACAAGATGGCGCGCATCAACG......ATTTGGCAGATATGGCGCAACTC 852
       |||    |||:::||||||        |||                  :::
 289 ArgArgHisArgThrSerThrGlyHisIleProProThrThrThrSerAr 305

 853 AAAGACTATGCCG...........CAGCAGCCATCCGCGATTGGGCAGT 890
         |||    |||             |||:::|||       :::    ||||
 305 gProThrGlyProProSerArgLeuGlnArgProValHisLeuTyrGlnS 322

 891 CCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATA...... 934
```

```
        ||:::           |||:::          :::||||||:::|||
 322 erSerProHisThrProAsnPhe......ArgProSerSerIleArgLys 336

 935 ...TCTTTATGGCAGCCATCCCCATCAAAGGGA...TTGGAGCTGTCCGG 978
         :::|||   |||   :::|||      |||   ||||||     ||
 337 AspAlaLeuLeuGlnThrGlyProArgLeuGlyHisLeuGluCysLeuGl 353

 979 GGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCA 1028
     |:::   :::          :::|||:::::        :::    ||||
 353 yGlnProAlaAsnLeuArgThrSerGluArgSerProProThrLysArgA 370

1029 GATGGGCGCGATCGCATTGCCGAAAGGGAAATCCG.........CCGTCA 1069
      ||        ||||||        :::      |||      |||:::
 370 rgLeuProArgSerSerGluProAsnArgLeuProLysProLeuProGlu 386

1070 GCGACAATTTTGCCG...ATGCGGCAT.................ACGCC 1098
     ||||||:::   |||      ||||||                    ||
 387 AlaThrLeuAlaProSerTyrArgHisArgArgSerTyrProPheLeuPr 403

1099 AAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCA 1148
      ||||       |||    :::::::|||:::|||:::          ::::::|
 403 oAsnPro...ProAlaAlaLeuProSerIleAla.........TyrThrS 416

1149 GCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACG 1198
      ||    |||    :::    ||||||    |||    ::::::|||:::    ::
 416 er.SerArgGlyLysIleHisHisSerLeuProLysGlyAlaLeuProLy 432

1199 GCAAAAATGTCAAACTGGCAGACCAACGCCACCC 1232
      ::::               |||    |||    ||||||
 432 sGlu...........GlyAlaProProProPro 439
```

seq_name: SwissProt_40:YS8A_CAEEL

```
seq_documentation_block:
ID   YS8A_CAEEL     STANDARD;      PRT;   796 AA.
AC   Q09625;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 84.3 kDa protein ZK945.10 in chromosome II.
GN   ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z48544; CAA88444.1; -.
DR   EMBL; Z48582; CAA88444.1; JOINED.
DR   WormPep; ZK945.10; CE01732.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      11     30       POTENTIAL.
FT   DOMAIN       273    546       SER/THR-RICH.
FT   DOMAIN       656    752       SER/THR-RICH.
SQ   SEQUENCE   796 AA;  84306 MW;  76DC5B03E6357A6A CRC64;

alignment_scores:
          Quality:  124.00            Length:    395
            Ratio:  0.649               Gaps:     17
 Percent Similarity:  48.354   Percent Identity:  22.532
```

```
alignment_block:
 US-09-303-518D-463 x YS8A_CAEEL   ..

 Align seg 1/1  to: YS8A_CAEEL  from: 1  to: 796

 110 TCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGC 159
     ||||||:::|||   |||:::    :::   :::||||||   ::::::::
 329 SerThrAlaSerThrSerThrSerSerThrSerThrThrGlnGlnSerSe 345

 160 AGGG......GGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAA 203
     ::::            |||   ||||||:::|||::::::|||
 345 rSerThrIleThrSerSerProSerSerThrThrLeuSer.......... 358

 204 CATACAAAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTG 253
                :::|||     :::|||
 359 .....ThrSerIleProThrThrThr.....................Thr 366

 254 AAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGC........ 295
        |||||::::::||||||||::::::|||    ::::::
 367 ProGluIleThrSerThrLeuSerSerLeuProAspAsnAlaIleCysSe 383

 296 ................ACAAATTCCATTCGCCCTTCGACAACCATGCCTC 329
                     |||:::|||        :::|||||||||
 383 rTyrLeuAspGluThrThrThrSerThrThrPheThrThrThrMetLeuT 400

 330 ACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTT 379
                     :::|||    :::       |||    :::
 400 hrSerThrThrThrGluGluProSerThrSerThrThrThrThrGluVal 416

 380 ACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGAC 429
     |||::::::    :::|||    |||:::|||    |||    |||::::::||
 417 ThrSerThrSerSerThrValThrThrThrGluProThrThrThrLeuTh 433

 430 GGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATA 479
     |            |||:::                                  |
 433 rThrSerThrAlaSer...............................T 439

 480 CAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCG 529
     ||:::||||||                      :::    ||||||
 439 hrSerThrThr.....................GluProSerThrSerThr 448

 530 ACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGC 579
     :::|||::::::|||:::::::::     |||::::::||||||:::
 449 ValThrThrSerProSerThrSer...ProValThrSerThrValThr.. 463

 580 GCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATA 629
                          :::::::::|||::::::              |
 464 .....................SerSerSerSerSerSerThrThrValT 473

 630 CAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCA 679
     ||:::|||:::   |||:::    :::              ||||||:::
 473 hrThrProThrSerThrGluSerThrSerThrSerProSerSerThrVal 489

 680 CTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGC 729
               ::::::    |||::::::               |||::
 490 ThrThrSerThrThrAlaProSerThrSerThrThrGlyProSerSerSe 506

 730 GCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCAT 779
     ::::::    |||                :::    |||::::::    :::|||
 506 rSerSerThrProSer.........SerThrAlaSerSerSerValSerS 520

 780 GCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACG 829
       ||||||                                    :::||||||
 520 erThrAla...............................SerSerThr 525

 830 ATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGC 879
                        :::|||:::|||      ||||||    |||::
 526 ..................GlnSerSerThrSerThrGlnGlnSerSerTh 536

 880 GATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCG.... 925
     :        :::|||:::|||                 :::    :::|||
 536 rThrThrLysSerGluThrThrThrSerSerAspGlyThrAsnProAspP 553
```

```
 926 .............TCAGCAATATCTTTATGGCAGCCATCCCCATCAAAG 961
                   :::|||   :::   :::     |||   |||
 553 heTyrPheValGluLysAlaThrThrThrPheTyrAspSerThrSerVal 569

 962 GGATTGGAGCTGTCCGGG............................... 979
     :::|||   |||::::::
 570 AsnLeuThrLeuAsnSerGlyLeuGlyIleIleGlyTyrGlnThrSerIl 586

 980 .GAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGC......G 1022
      |||   |||:::   ::::::|||:::::::::   ::::::     |
 586 eGluCysThrSerProThrSerSerAsnTyrValSerThrThrLysAspG 603

1023 GTCGCAGATGGGCGCGATCGCATTGCCGAAAG.........GGAAATCCG 1063
      ||      :::::::::|||                    |||:::
 603 lyAlaCysPheThrLysSerValSerMetProArgLeuGlyGlyThrTyr 619

1064 CCGTCAGCGACAATTTTG...............CCGATGCGGCATACGCC 1098
      |||::::::|||::::::                   :::|||   |||
 620 ProAlaSerThrPheValGlyProGlyAsnTyrThrPheArgAlaThrMe 636

1099 AAATACCCGTCCC............CTTACCATTCCCGAAATATCCGTTC 1136
      :::|||                   |||      ::::::   :::|
 636 tThrThrAspAspLysLysValTyrTyrThrTyrAlaAsnValTyrIleG 653

1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCA 1171
      ||   :::|||||||:::|||    ::::::   |||
 653 lnGluTyrSerSerThrThrIleGluSerGluSer 664
```

seq_name: SwissProt_40:VGP3_EBV

seq_documentation_block:

```
ID   VGP3_EBV       STANDARD;      PRT;   907 AA.
AC   P03200; P03201;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
DE   Glycoprotein GP220].
GN   BLLF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
CC   -!- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC       B-CELLS.
CC   -!- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC       ENVELOPE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; V01555; CAA24854.1; -.
DR   PIR; A03762; QQBE21.
DR   PIR; A03763; QQBE22.
DR   PIR; S33008; S33008.
KW   Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    277    277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    443    443       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    497    497       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    568    568       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    610    610       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    627    627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    683    683       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    701    701       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    735    735       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    746    746       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    755    755       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    780    780       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    815    815       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    858    858       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    888    888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    502    698       MISSING (IN GP220).
SQ   SEQUENCE   907 AA;  94431 MW;  0750141CBCAC52C9 CRC64;
```

alignment_scores:

```
          Quality:  123.50          Length:    412
            Ratio:  0.647             Gaps:     17
 Percent Similarity:  46.359   Percent Identity:  22.330
```

alignment_block:
 US-09-303-518D-463 x VGP3_EBV ..

 Align seg 1/1 to: VGP3_EBV from: 1 to: 907

```
 110 TCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGC 159
     ||||||::::::::::::::::||||||            |||   |||:::||
 469 SerThrAlaAspValThrSerProThrProAlaGlyThrThrSerGlyAl 485

 160 AGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACA 209
     |:::              ||||||               |||   |||:::
 485 aSerProValThrProSerPro.........SerProTrpAsp....... 496

 210 AAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAA 259
                                  |||            |||
 497 ...........................AsnGlyThrGluSerLysAla 503

 260 ATATCGGCTACATTGTCCGCTTTTCCGATCACG...........GGCAC 297
        :::   |||   |||:::   |||:::|||               :::::
 504 ProAspMetThrSerSerThrSerProValThrThrProThrProAsnAl 520

 298 AAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGA 347
     ::::|||         |||::::::|||   |||      |||     |||:
 520 aThrSerProThrProAlaValThrThrPro...ThrProAsnAlaThrS 536

 348 AGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACG 397
     ::|||:::        :::|||      :::   :::|||:::       |||
 536 erProThrProAlaValThrThrProThrProAsnAlaThrSerProThr 552
```

```
 398 GATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGC 447
               :::|||   |||   :::|||:::|||
 553 LeuGlyLysThrSerProThrSerAlaValThr................. 563

 448 TATCCCGCTCCCAAAGGCG............................... 466
        |||   |||:::|||
 564 .ThrProThrProAsnAlaThrSerProThrLeuGlyLysThrSerProT 580

 467 ..CGAGGGATATATACAGCTACG.........ACATAAAAGGCGTTGCCC 505
         ::::::   |||   |||           |||     :::|||
 580 hrSerAlaValThrThrProThrProAsnAlaThrSerProThrLeuGly 596

 506 AAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTT 555
     |||   |||   ::::::       ||||||  :::|||:::
 597 LysThrSerProThrSerAlaValThrThrProThrProAsnAlaThrGl 613

 556 GCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGA 605
      |||||||||:::
 613 yProThrValGly..................................... 617

 606 CGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCA 655
          :::::::::|||      :::|||   :::   |||       :::
 618 .....GluThrSerProGlnAlaAsnAlaThrAsnHisThrLeuGlyGly 632

 656 ATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATC 705
         |||   |||   ::::::   |||   ::::::   ||||||::
 633 ThrSerProThrProValValThrSerGlnProLysAsnAlaThrSerAl 649

 706 GGCGCGG......CAGGAGAAATTG.........TCGGCGCAGGCGATGC 740
     ::::         |||:::::::::          |||:::::::::|||
 649 aValThrThrGlyGlnHisAsnIleThrSerSerSerThrSerSerMetS 666

 741 CGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGG 790
        |||       ::::::   :::||||||        |||:::
 666 erLeuArgProSerSerAsnProGluThrLeuSerProSerThrSer... 681

 791 GTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGAT 840

 681 .................................................. 681

 841 ATGGCGCAACTCAAAGACT...ATGCCGCAGCAGCCATCCGCGATT.... 883
            ||||||   :::   ||||||       ||||||
 682 ....AspAsnSerThrSerHisMetProLeuLeuThrSerAlaHisProT 697

 884 .....GGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCA 928
          |||::::::   |||   :::         :::   :::
 697 hrGlyGlyGluAsnIleThrGlnValThrProAlaSerIleSerThrHis 713

 929 GCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGG 978
        :::|||      :::|||:::|||      |||
 714 HisValSerThrSerSerProAlaProArgProGly.............. 725

 979 GGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCA 1028
        :::|||:::   |||::::::              |||::::::
 726 ....ThrThrSerGlnAlaSerGlyProGlyAsnSerSerThrSerThrL 741

1029 GATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATT 1078
     ::   ::::::   :::      ||||||:::||||||:::   :::
 741 ysProGlyGluValAsnValThrLysGlyThrProProGlnAsnAlaThr 757

1079 TTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAAT 1128
       |||       ::::::|||:::   :::      ::::::|||   :::
 758 SerPro......GlnAlaProSerGlyGlnLysThrAlaValProThrVa 772

1129 ATCCGTTCAA......ACTTGGAGCAGCGTTACGGCAAAGAAAACATCAC 1172
     ::::                   :::|||:::|||:::   |||   :::
 772 lThrSerThrGlyGlyLysAlaAsnSerThrThrGlyGlyLysHisThrT 789

1173 CTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACC 1222
        :::   :::|||   :::|||         :::::::::
 789 hrGlyHisGlyAlaArgThrSerThrGluProThrThrAspTyrGlyGly 805

1223 AACGCCACCCGA...AGACAGGCGTACCGTTTGACG 1255
```

```
        ::::::|||       |||       |||:::   |||
    806 AspSerThrThrProArgProArgTyrAsnAlaThr 817
```

seq_name: SwissProt_40:EGR1_BRARE

```
seq_documentation_block:
ID   EGR1_BRARE     STANDARD;      PRT;   511 AA.
AC   P26632;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Early growth response protein 1 (EGR-1) (Krox24).
GN   EGR1.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC   Cypriniformes; Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95032735; PubMed=7945937;
RA   Drummond I.A., Rohwer-Nutter P., Sukhatme V.P.;
RT   "The zebrafish egr1 gene encodes a highly conserved, zinc-finger
RT   transcriptional regulator.";
RL   DNA Cell Biol. 13:1047-1055(1994).
RN   [2]
RP   SEQUENCE OF 315-376 FROM N.A.
RX   MEDLINE=92028854; PubMed=1930167;
RA   Lanfear J., Jowett T., Holland P.W.;
RT   "Cloning of fish zinc-finger genes related to Krox-20 and Krox-24.";
RL   Biochem. Biophys. Res. Commun. 179:1220-1224(1991).
CC   -!- FUNCTION: TRANSCRIPTIONAL REGULATOR. RECOGNIZE AND BINDS TO THE
CC       DNA SEQUENCE 5'-CGCCCCCGC-3'(EGR-SITE). ACTIVATE THE TRANSCRIPTION
CC       OF TARGET GENES WHOSE PRODUCTS ARE REQUIRED FOR MITOGENESIS AND
CC       DIFFERENTIATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- INDUCTION: BY GROWTH FACTORS.
CC   -!- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U12895; AAA63651.1; -.
DR   EMBL; M81109; AAA50035.1; -.
DR   PIR; PQ0233; PQ0233.
DR   HSSP; P08046; 1AAY.
DR   ZFIN; ZDB-GENE-980526-320; egr1.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 3.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 3.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Repeat; Zinc-finger; Metal-binding.
FT   DOMAIN      146    152       POLY-SER.
FT   DOMAIN      155    165       POLY-SER.
FT   DOMAIN      311    391       ZINC FINGERS.
FT   ZN_FING     311    335       C2H2-TYPE.
FT   ZN_FING     341    363       C2H2-TYPE.
FT   ZN_FING     369    391       C2H2-TYPE.
FT   CONFLICT    317    317       T -> S (IN REF. 2).
FT   CONFLICT    320    320       R -> S (IN REF. 2).
FT   CONFLICT    351    351       S -> R (IN REF. 2).
FT   CONFLICT    372    372       E -> D (IN REF. 2).
SQ   SEQUENCE   511 AA;  55139 MW;  411C31E6F4AA10BF CRC64;
```

```
alignment_scores:
           Quality:  123.00            Length:    423
             Ratio:  0.597               Gaps:     30
 Percent Similarity:  48.700   Percent Identity: 24.823

alignment_block:
 US-09-303-518D-463 x EGR1_BRARE   ..

 Align seg 1/1  to: EGR1_BRARE  from: 1  to: 511

     15 CAAAATATCCCTTATTCTGTCCATACTGGCAGTGTG.........CCTGC 55
        |||:::::|||          :::||||||             |||||
     90 GlnArgLeuProProIleSerTyrThrGlyArgPheThrLeuGluProAl 106

     56 CGATGCATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAG 105
         |       |||::::::  ::::::|||       ||||||
    106 aThrAsnCysSerAsn..SerLeuTrpAlaGlu...ProLeu........ 118

    106 GTTCTCGACCGTC................................AGCA 122
          ||||||  |||                                ||||
    119 .PheSerLeuValSerGlyLeuValGlyIleAsnProProProAlaSerI 135

    123 TTTCGAACCCGACGGGAAATACCACCTAT......TCGGCAGCAGGGGGG 166
        ||    :::    |||:::::::::|||::::    |||:::::   :::
    135 leProSerSerThrSerGlnAlaThrHisProSerSerSerSerThrSer 151

    167 AGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCCAT 216
        |||:::||||||:::::::::    |||          :::     :::::
    152 SerIleProSerSerSerSerSerSer......ThrSerSerAlaSerLe 166

    217 CAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGG 266
         :|||   ::::::         :::   :::|||:::   :::   :::
    166 uSerCysSerValHisGlnSerGluProAsnProIleTyrSerAlaAlaP 183

    267 CTACATTGTCCGCT...............TTTCCGATCACGGGCACAAAT 301
           |||    |||:::               ||||||   :::|||   |||
    183 roThrTyrSerSerAlaSerProAspIlePheProGluSerGlyProAsn 199

    302 .....................TCCATTCGC...CCTTCGACAACCATGCC 327
                             |||::::::      |||:::|||   ||
    200 PheSerThrThrValGlyThrSerLeuGlnTyrSerSerSerThrTyrPr 216

    328 TCACATT...........CCGATTCTGACGAAGCCGGTAGTCCCGTTGA 365
        |                   |||   ::::::   |||::::::|||
    216 oSerAlaLysThrCysAsnProSerPheSerValProMetIlePro.... 231

    366 CGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCG 415
           |||      ||||||   :::        ::::::   :::    :::|||
    232 ..AspTyrLeuPheThrGlnGln...GlnSerGluIleSerLeuValPro 246

    416 CCG...............ACGGCTATGACGGGCCACAGGGCGGCGGCTAT 450
        |||                ::::::::|||::::::    |||
    247 ProAspGlnLysProIleGlnThrGlnAlaGlyGlnGlnProAlaLeuTh 263

    451 CCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGG... 497
         ||||||            :::::::::   |||    |||      ::::::
    263 rProLeuHisThrIleLysAlaPheAlaThrGlnThr.GlySerGlnAsp 279

    498 ......................................CGTTGCCCAAA 508
                                               |||    ||
    280 LeuLysSerValTyrGlnSerGlnLeuIleLysProSerArgMetArgLy 296

    509 ATATCCGCCTCAACCTGACCGACAACCGCAGCAC...CGGACAACGGCTT 555
         ||||||    :::|||::::::   |||    |||    |||      ||||
    296 sTyrProAsnArgProSerLysThrProProHisGluArgProTyrAlaC 313

    556 GCCGACCGTTTCCACAATGC...CGGCGCTATGCTGACGCAAGGAGTAGG 602
         ||    |||          |||   ||||||:::            ||||||
    313 ys...ProValGluThrCysAspArgArgPhe............SerArg 324

    603 CGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGG 652
                  |||||||||   |||:::    ::::::||||||
    325 SerAspGluLeuThrArgHis...IleArgIleHisThrGlyGln..... 338

    653 GCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATC 702
                  :::   :::|||        ||||||      :::::::::
    339 ..........LysProPheGln......CysArgIleCysMetArgAsnP 350

    703 ATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTAT 752
              |||:::   :::            :::   |||   ::::::|||
    350 heSerArgSerAspHisLeuThrThrHisIleArgThrHisThrGlyGlu 366

    753 AAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCA 802
         |||             :::    |||       :::::::::    :::    ::
    367 LysProPheAlaCysGluIleCysGlyArgLysPheAlaArgSerAspGl 383

    803 CCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTC 852
        :||||||  :::    ::::::|||:::              |||    |||
    383 uArgLys.ArgHisThrLysIleHis............MetArgGln... 394

    853 AAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAA 902
         ||||||       |||                     |||||||||:::
    395 LysAspLysLysAlaGluLysGlyAlaThrAlaAlaValGlnSer..... 409

    903 TGCCGCACAAGGCATAGAAGCCGTCAGCAATATC...TTTATGGCAGCCA 949
                                  :::||||||||||||    :::    |||:::
    410 ..................SerValSerAsnIleSerIleSerAlaSerS 420

    950 TCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGC 999
            |||:::::::::         :::
    420 erProValSerSerTyrProSer......................Pro 428

   1000 ATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCC 1049
         ||||||::::::|||          |||::::::
    429 IleThrSerTyrProSerProValSerSerPhe................ 439

   1050 GAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCA 1099
                     |||    |||::::::   :::::::::    :::::::::
    440 .......ProSerProValAsnSerCysTyrSerSerProValHisThrS 454

   1100 AATACCCGTCCCCT 1113
        ::||||||||||||
    454 erTyrProSerPro 458

seq_name: SwissProt_40:A180_MOUSE

seq_documentation_block:
ID    A180_MOUSE     STANDARD;      PRT;   901 AA.
AC    Q61548; Q61547;
DT    01-NOV-1997 (Rel. 35, Created)
DT    01-NOV-1997 (Rel. 35, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    Clathrin coat assembly protein AP180 (Clathrin coat associated protein
DE    AP180) (Phosphoprotein F1-20) (91 kDa synaptosomal-associated
DE    protein).
GN    SNAP91.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=92300439; PubMed=1607933;
RA    Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
RT    "Characterization of a novel synapse-specific protein. II. cDNA
RT    cloning and sequence analysis of the F1-20 protein.";
RL    J. Neurosci. 12:2144-2155(1992).
CC    -!- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES WHICH
CC        LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-
CC        ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE
CC        CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION
CC        AND CONCENTRATION. BINDING OF AP180 TO CLATHRIN TRISKELIA INDUCES
CC        THEIR ASSEMBLY INTO 60-70 NM COATS.
CC    -!- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
CC        CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.
```

```
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: BRAIN. ASSOCIATED WITH THE SYNAPSES.
CC   -!- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED IN A PATTERN
CC       COINCIDENT WITH ACTIVE SYNAPTOGENESIS AND SYNAPTIC MATURATION.
CC   -!- DOMAIN: POSSESSES A THREE DOMAIN STRUCTURE: THE N-TERMINAL 300
CC       RESIDUES HARBOR A CLATHRIN BINDING SITE, AN ACIDIC MIDDLE DOMAIN
CC       450 RESIDUES, INTERRUPTED BY AN ALA-RICH SEGMENT, AND THE C-
CC       TERMINAL DOMAIN (166 RESIDUES).
CC   -!- PTM: PHOSPHORYLATED.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M83985; AAA37587.1; -.
DR   EMBL; M83985; AAA37586.1; -.
DR   HSSP; P04002; 1WFA.
DR   MGD; MGI:109132; Snap91.
DR   InterPro; IPR001026; ENTH.
DR   Pfam; PF01417; ENTH; 1.
DR   SMART; SM00273; ENTH; 1.
KW   Coated pits; Alternative splicing; Phosphorylation.
FT   DOMAIN      410    413       POLY-THR.
FT   DOMAIN      535    539       POLY-ALA.
FT   DOMAIN      547    550       POLY-ALA.
FT   DOMAIN      659    664       POLY-SER.
FT   DOMAIN      704    710       POLY-SER.
FT   VARSPLIC    715    719       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   901 AA;  91851 MW;  24A98FBACE8DB8B1 CRC64;

alignment_scores:
           Quality:  123.00             Length:     552
             Ratio:  0.572                Gaps:      27
 Percent Similarity:  38.949   Percent Identity:  22.101

alignment_block:
 US-09-303-518D-463 x A180_MOUSE   ..

 Align seg 1/1  to: A180_MOUSE  from: 1  to: 901

      92 CCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAA 141
          |||:::|||      |||:::      :::    |||   ||||||
     397 ProIleSerAspProPheAlaPro...GluProSerProProThr..... 410

     142 TACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATAT 191
           |||||    :::    |||::::::|||       ::::::||||||::::
     411 .ThrThrThrGluProAlaSerAlaSerAlaSerThrThrThrAlaValT 427

     192 CGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAAC 241
          ::         |||
     427 hrAlaValThrThr.................................... 431

     242 AGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTT........ 283
                          |||:::::        :::||||||
     432 ..............GluValAspLeuPheGlyAspAlaPheAlaAlaSe 443

     284 .............CCGATCACGGGCACAAATTCCATTCGCCCTTCGACAA 320
                      |||    ::::::    ::::::      |||:::|||
     443 rProGlyGluAlaProAlaAlaSerGluGlyAlaThrAlaProAlaThrP 460

     321 CCATGCCTCACATTCCGATTCTG..............ACGAAGCCGGTA 355
              |||             |||              :::   |||
     460 roAlaProValAlaAlaAlaLeuAspAlaCysSerGlyAsnAspProPhe 476

     356 GTCCCGTTGACGGATTCAGCCTTTACCGCA.................... 385
          :::|||         ||||||   :::|||
     477 AlaProSerGluGlySerAlaGluAlaAlaProGluLeuAspLeuPheAl 493

     386 .......TCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGA 428
                          |||::::::  ::::::  |||||||||   |
     493 aMetLysProProGluThrSerAlaProValValThrProThrAlaSerT 510

     429 CGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATAT 478
          ||             |||   |||:::|||
     510 hr..............AlaProProValPro.................. 515

     479 ACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACC 528
            ||||||:::   :::        |||          |||:::|||
     516 ...AlaThrAlaProSerProAlaProThrAlaValAlaAlaThrAlaAl 531

     529 GACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGG 578
           ||||||:::|||             :::     |||::::::|||      |
     531 aThrThrThrAlaAlaAlaAlaAlaThrThrThrAlaThrThrSerAlaA 548

     579 CGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCA...... 622
          ||                  ||||||   :::   ||||||
     548 laAla..................AlaThrThrAlaAlaAlaProProAla 558

     622 .................................................. 622

     559 LeuAspIlePheGlyAspLeuPheAspSerAlaProGluValAlaAlaAl 575

     623 .......CCCGATACAGCCCCGAGC...........TGGACAGAT.... 649
                ||||||:::|||||||||               ||||||
     575 aProLysProAspAlaAlaProSerIleAspLeuPheGlyThrAspAlaP 592

     650 .............CGGGCAATGCCGCCGAAGCCTTCAACGGC....ACT 681
                      |||:::      |||    |||::::::       |||
     592 heSerSerProProArgGlyAlaSerProValProGluSerSerLeuThr 608

     682 GCAGATATCGTC...................................... 693
          ||||||::::::
     609 AlaAspLeuLeuSerValAspAlaPheAlaAlaProSerProAlaSerTh 625

     694 ........................AAAAACATCATCGGCGCGGCAGGAG 718
                                  :::::::::|||       |||:
     625 rAlaSerProAlaLysAlaGluSerSerGlyValIleAspLeuPheGlyA 642

     719 AAATTGTCGGCGCAGGC..............GATGCCGTGCAGGGTATA 753
          ::         |||:::|||              |||   |||::::::
     642 spAlaPheGlySerGlyAlaSerGluThrGlnProAlaProGlnAlaVal 658

     754 AGCGAAGGCTCAAACATTGCT...GTCATGCACGGCTTGGGT........ 792
          |||::::::|||      |||    ::::::   |||:::|||
     659 SerSerSerSerAlaSerAlaAspLeuLeuAlaGlyPheGlyGlySerPh 675

     793 ...........................................CTGCTTT 799
                                                   ||||||:
     675 eMetAlaProSerThrThrProValThrProAlaGlnAsnAsnLeuLeuG 692

     800 CCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAA 849
          :: |||:::   :::         :::|||      :::          ::
     692 ln.ProSerPheGluAlaAlaPheGlyThrThrProSerThrSerSerSe 708

     850 CTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCC 899
          :|||:::       |||:::      :::::::::|||:::         |
     708 rSerSerPheAspProSerValPheAspGlyLeuGlyAspLeuLeuMetP 725

     900 CAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCA 949
          ||            :::   :::|||:::   :::|||:::       |||
     725 roThrMetAlaProSerGlyGlnProAlaProValSerMetValProPro 741

     950 TCCCCA...........TCAAAGGGA........................ 964
          ||||||            |||||||||
     742 SerProAlaMetAlaAlaSerLysGlyLeuGlySerAspLeuAspSerSe 758

     965 ......................TTGGAGCTGTCCGGGGAAAATACG.... 988
                                |||    :::||||||    ::::::
     758 rLeuAlaSerLeuValGlyAsnLeuGlyIleSerGlyThrThrSerLysL 775
```

```
 989 ..........GCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGC 1027
                |||  |||:::::     |||:::::|||
 775 ysGlyAspLeuGlnTrpAsnAlaGlyGluLysLysLeuThrGlyGlyAla 791

1028 AGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAAT 1077
     :::|||
 792 AsnTrpGln......................................... 794

1078 TTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAA 1127
                                               |||:::
 795 ..........................................ProLysV 797

1128 TATCCGTTCAAACTTGGAGCAGC...GTTACGGCAAAGAAAACATCACCT 1174
     :::::      |||||||||:::  |||     :::  |||  |||
 797 alThrProAlaThrTrpSerAlaGlyValProProGlnGlyThrValPro 813

1175 CCTCA....ACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGA 1220
     |||      :::|||||||||:::  |||
 814 ProThrSerSerValProProGlyAlaGlyAla................. 824

1221 CCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGA 1270
                 |||::::::|||  |||     |||  |||:::|||
 825 ..........ProSerValGlyGlnProGlyAlaGlyPheGlyMetProP 838

1271 ATTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTA 1320

 838 ro............................................... 838

1321 TCGGGGGGCGGTATACCT......AAGGCTAAGCCTGTGTTTGATGCGAA 1364
     ||||||  |||:::          :::  :::|||||||:::  |||::
 839 SerGlyThrGlyMetThrMetMetSerGlnGlnProValMetPheAlaGl 855

1365 ACCG 1368
      :|||
 855 nPro 856
```

seq_name: SwissProt_40:MGPC_MYCGE

```
seq_documentation_block:
ID   MGPC_MYCGE     STANDARD;      PRT;  1052 AA.
AC   P22747; Q49481; Q49482; Q49284; Q49385; Q49257;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   MgpC protein precursor.
GN   MGPC OR MG192.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=90060815; PubMed=2583522;
RA   Inamine J.M., Loechel S., Collier A.M., Barile M.F., Hu P.-C.;
RT   "Nucleotide sequence of the MgPa (mgp) operon of Mycoplasma
RT   genitalium and comparison to the P1 (mpp) operon of Mycoplasma
RT   pneumoniae.";
RL   Gene 82:259-267(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995):
RN   [3]
RP   SEQUENCE OF 60-133; 260-370; 441-512 AND 964-1052.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
RN   [4]
RP   SEQUENCE OF 769-964 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=92051396; PubMed=1945886;
RA   Peterson S.N., Schramm N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A random sequencing approach for placing markers on the physical map
RT   of Mycoplasma genitalium.";
RL   Nucleic Acids Res. 19:6027-6031(1991).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M31431; AAA25421.1; -.
DR   EMBL; U39698; AAC71411.1; -.
DR   EMBL; U02124; AAD12399.1; -.
DR   EMBL; U02161; AAD12443.1; -.
DR   EMBL; U34967; AAA88890.1; -.
DR   EMBL; U02157; AAD12439.1; -.
DR   EMBL; U34970; AAA88899.1; -.
DR   EMBL; X61525; CAB98130.1; -.
DR   EMBL; X61527; CAB98131.1; -.
DR   PIR; JQ0092; JQ0092.
DR   TIGR; MG192; -.
KW   Cell adhesion; Signal; Membrane; Complete proteome.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1052       MGPC PROTEIN.
FT   CONFLICT     60     68       LDSSYQIAD -> MIKPLPLLS (IN REF. 3).
FT   CONFLICT    125    126       TM -> SV (IN REF. 3).
FT   CONFLICT    474    481       TNNGTTGP -> SNQISSGT (IN REF. 3;
FT                                AAD12439).
FT   CONFLICT    476    476       N -> S (IN REF. 3; AAA88890).
FT   CONFLICT    897    902       SVSPRI -> MSHQGS (IN REF. 4).
SQ   SEQUENCE   1052 AA;  114360 MW;  F6C6D3B3D7789145 CRC64;
```

```
alignment_scores:
           Quality:  123.00           Length:    686
             Ratio:  0.444              Gaps:     30
 Percent Similarity:  40.379   Percent Identity: 17.930

alignment_block:
 US-09-303-518D-463 x MGPC_MYCGE   ..

 Align seg 1/1  to: MGPC_MYCGE  from: 1  to: 1052

 151 TTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATC...GGATT 197
      ::::::|||  ::::::||||||      |||:::||||||     ::
 296 TyrAlaSerThrSerLysLeuAlaValThrAsnAsnHisIleValValMe 312

 198 GGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGG 247
      :||||||            ::::::  :::  ::::::::::::::|
 312 tGlyAsnSerPheLeuProSerMetTrpTyrTrpValValGluArgSerA 329

 248 CC................................................ 249
      ||
 329 laGlnGluAsnAlaSerAsnLysProThrTrpPheAlaAsnThrAsnLeu 345

 250 .............................GTTGAAGGAAATATCGGCTA 269
                                  ||||||::::::::|||||
 346 AspTrpGlyGluAspLysGlnLysGlnPheValGluAsnGlnLeuGlyTy 362

 270 CATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCC....... 312
```

```
       |             |||   :::::|||:::||||||||
362 rLysGluThrThrSerThrAsnSerHisAsnPheHisSerLysSerPheT 379

312 .................................................. 312

379 hrGlnProAlaTyrLeuIleSerGlyIleAspSerValAsnAspGlnIle 395

313 .................................TTCGACAACCATGCCTC 329
                                     :::|||:::   :::||
396 IlePheSerGlyPheLysAlaGlySerValGlyTyrAspSerSerSerSe 412

330 ACATTCCGATTCT..............GACGAAGCC............. 351
    |   |||:::|||              |||:::|||
412 rSerSerSerSerSerSerSerThrLysAspGlnAlaLeuAlaTrpSerT 429

351 .................................................. 351

429 hrThrThrSerLeuAspSerLysThrGlyTyrLysAspLeuValThrAsn 445

352 ...........GGTAGTCCCGTTGACGGA...TTCAGCCTT........ 378
               ::::::|||:::::|||   |||||||:::
446 AspThrGlyLeuAsnGlyProIleAsnGlySerPheSerIleGlnAspTh 462

379 .......TACCGCATCCATTGGACGGATACGAACACCATCCCGCCGACG 421
           :::    :::   ::::::|||          :::|||        :::|
462 rPheSerPheValValProTyrSerGly......AsnHisThrAsnAsnG 477

422 GCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAA......... 462
    ||        ||||||         :::||||||:::   |||
477 lyThrThrGlyProIleLysThrAlaTyrProValLysLysAspGlnLys 493

463 ..........................GGCGCGAGGGATATATACAGCTA 485
                              :::|||      :::   |||||
494 SerThrValLysIleAsnSerLeuIleAsnAlaThrProLeuAsnSerTy 510

486 CGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAAC. 534
    |       :::|||::::::              |||:::|||:::   |||
510 rGlyAspGluGlyIleGlyValPheAspAlaLeuGlyLeuAsnTyrAsnP 527

535 ..CGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCT 582
       :::|||:::   :::||||||   :::|||          :::
527 heLysSerAsnGlnGluArgLeuProSerArg.........ThrAspGln 540

583 ATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAG 632
    :::::::::   |||:::                          ||
541 IlePheValTyrGlyIle..........................ValSe 548

633 CCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTG 682
    ||||:::      ||||||::::::::::::         :::|||   :
548 rProAsnGluLeuArgSerAlaLysSerSerAlaAspSerThrGly...S 564

683 CAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 732
    ::|||         |||      ::::::   ::::::   :::
564 erAspThrLysValAsnTrpSerAsnThrGlnSerArgTyrLeu...... 578

733 GGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCA 782
          :::      :::   |||||||||            ::::::
579 ...ProValProTyrAsnTyrSerGluGly...........IleIleAs 590

783 CGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATT 832
     :::    |||:::          ||||||:::   |||   ::::::   :
590 pAlaAspGlyPheLysArgProGluAsnArgGlyAlaSerValThrThrP 607

833 TGGCAGATATGGCGCAACTC...AAAGACTATGCCGCAGCAGCCATCCGC 879
    :::::   :::   ::::::         |||      |||   :::|||
607 heSerGlyLeuLysSerIleAlaProAspGlyPheAlaAsnSerIleAla 623

880 GATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAG 929
    :::::::::|||                      |||:::::::|||
624 AsnPheSerVal....................GlyLeuLysAla..... 631

930 CAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGG 979

631 .................................................. 631

980 GAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTC.........AAG 1020
                      ||||||    ::::||||||           |||
632 .................GlyIleAspProAsnProValMetSerGlyLys 642

1021 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGG.............. 1056
     :::::::::   ||||||::::::|||    :::|||
643 LysAlaAsnTyrGlyAlaValValLeuThrArgGlyGlyValValArgLe 659

1057 ..........................AAATCCGCCGTCAGCGACAATT 1078
                               |||:::::::::::::|||:
659 uAsnPheAsnProGlyAsnAspSerLeuLeuSerThrThrAspAsnAsnI 676

1079 TTGCCGATGCGGCATACGCCAAATACCCG......TCCCCTTACCATTCC 1122
     ::|||      :::::::::        |||         :::            :::
676 leAlaProIleSerPheSerPheThrProPheThrAlaAlaGluSerAla 692

1123 CGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAAC..... 1167
       ::::::   :::         |||      |||::::::|||:::
693 ValAspLeuThrThrPheLysGluValThrTyrAsnGlnGluSerGlyLe 709

1168 ..........ATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATG 1207
               |||    :::|||::::::   ||||||:::
709 uTrpSerTyrIlePheAspSerSerLeuLysProSerHis.......... 722

1208 TCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGT 1257
                                                    ||||||
723 ...........................................AspGly 724

1258 AAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGA 1307
     |||       |||                |||       :::      :::::
725 LysGlnThrPro..............ValThrAspAsnMetGlyPheSe 736

1308 TATTCAAGAATTATCGGGGGGCGGTATA...........CCTAAGGCTA 1345
     ::::       :::|||      ||||||                  :::|||
736 rValIleThrValSerArgThrGlyIleGluLeuAsnGlnAspGlnAlaT 753

1346 AGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTTAAT 1395
         ::::::|||:::   |||
753 hrThrThrLeuAspValAlaPro........................... 760

1396 AAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAG 1445
                   :::          |||:::::::::::|||:::|||    ::
761 .........SerAlaLeuAlaValGlnSerGlyIleGlnSerThrThrGl 774

1446 AAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAA 1495
     :
774 n................................................. 774

1496 ATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAG 1545

774 .................................................. 774

1546 AAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGTGATGTACGGGT 1595
           |||:::||||||        |||::::::                  ||
775 ......ThrLeuThrGlyValLeuProLeuSerGluGluPheSerAlaVa 789

1596 GATACAACAAACCTCGGCACCTGATAAACATGGGGTTTATCAAGCGACAG 1645
     ||||   :::   |||         :::|||      :::|||
789 lIleAlaLysAspSerAspGlnAsnLysIleAspIleTyr.......... 802

1646 TGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTGAAAACGAAAAAAGGT 1695
             |||:::   :::|||   :::|||:::   |||:::   :::
803 ........LysAsnAsnAsnGlyLeuPheGluIleAspThrGlnLeuSer 816

1696 GGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAGATTGGGATGAGGC 1745
     ::::::|||   |||::::::         :::      ::::::   |||
817 AsnSerValAlaThrAsnAsnGlyGlyLeuAlaProSerTyrThrGluAs 833

1746 TAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTA 1795
      |||:::                   :::||||||   :::::::::   :::
833 nArgVal..............AspAlaTrp...GlyLysValGluPheA 844
```

```
  1796 AGGATAAT 1803
         ||||||
   844 laAspAsn 846

seq_name: SwissProt_40:GTFC_STRMU

seq_documentation_block:
ID   GTFC_STRMU     STANDARD;      PRT;  1375 AA.
AC   P13470; P05427;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE   (Dextransucrase) (Sucrose 6-glucosyltransferase).
GN   GTFC.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GS-5;
RX   MEDLINE=89137980; PubMed=2976010;
RA   Ueda S., Shiroza T., Kuramitsu H.K.;
RT   "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL   Gene 69:101-109(1988).
RN   [2]
RP   SEQUENCE OF 1-349 FROM N.A.
RC   STRAIN=GS-5;
RX   MEDLINE=87308013; PubMed=3040685;
RA   Shiroza T., Ueda S., Kuramitsu H.K.;
RT   "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL   J. Bacteriol. 169:4263-4270(1987).
CC   -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC       TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC       OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC       AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC   -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
CC       fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DISEASE: DENTAL CARIES.
CC   -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC       1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC       WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC       FORMS OF GLUCANS.
CC   -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC       BINDING PROTEIN FROM S.MUTANS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M22054; AAA88592.1; -.
DR   EMBL; M17361; AAA88589.1; -.
DR   PIR; JT0345; JT0345.
DR   PIR; C33135; C33135.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR003318; Glyco_hydro_70.
DR   Pfam; PF01473; CW_binding_1; 7.
DR   Pfam; PF02324; Glyco_hydro_70; 1.
KW   Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT   SIGNAL        1     34
FT   CHAIN        35   1375       GLUCOSYLTRANSFERASE-SI.
FT   DOMAIN       35   1050       CATALYTIC (APPROXIMATE).
FT   DOMAIN     1126   1375       GLUCAN-BINDING (APPROXIMATE).
FT   DOMAIN     1126   1375       2.4 A, 1 C AND 1 AC REPEATS.
FT   REPEAT     1126   1159       A REPEAT.
FT   REPEAT     1169   1200       A REPEAT.
FT   REPEAT     1227   1238       C REPEAT.
FT   REPEAT     1253   1303       AC REPEAT.
FT   REPEAT     1318   1330       A REPEAT (INCOMPLETE).
SQ   SEQUENCE   1375 AA;  153022 MW;  D4B80CBEE0AACE13 CRC64;

alignment_scores:
            Quality:  123.00              Length:    533
              Ratio:  0.557                 Gaps:     32
 Percent Similarity:  41.463    Percent Identity: 22.889

alignment_block:
 US-09-303-518D-463 x GTFC_STRMU   ..

 Align seg 1/1  to: GTFC_STRMU  from: 1  to: 1375

       64 GCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGA 113
           :::::|||      |||:::|||          :::    :::|||
      643 SerTyrAlaLeuLeuLeuThrAsnLysSerSerValProArgValTyrTy 659

      114 CCGTCAGCATTTCGAACCCGACGGGAAATAC................... 144
             :::    |||      |||||||:::|||
      659 rGlyAspMetPheThrAspAspGlyGlnTyrMetAlaHisLysThrIleA 676

      145 .................CACCTATTCGGCAGCAGG............... 162
                            |||:::    :::|||
      676 snTyrGluAlaIleGluThrLeuLeuLysAlaArgIleLysTyrValSer 692

      163 GGGGAGCTTGCCNAGCGCAACGGCCATATC.................... 192
           |||       |||    ||||||    ::::::
      693 GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGluIleIleTh 709

      193 .............GGATTGGGAAACATACAAAGCCATCAGTTGGGCCACC 229
                        |||    |||    :::::::::    :::    |||
      709 rSerValArgTyrGlyLysGlyAlaLeuLysAlaThrAspThrGlyAspA 726

      230 TGATGATTCAACAG.........GCGGCCGTTGAAGGAAATATCGGCTAC 270
                :::               |||::::::|||||||||
      726 rgThrThrArgThrSerGlyValAlaValIleGluGlyAsnAsnProSer 742

      271 ATT...GTCCGCTTTTCCGAT.........................CA 290
           :::    ::::::    ||||||                          ||
      743 LeuArgLeuLysAlaSerAspArgValValValAsnMetGlyAlaAlaHi 759

      291 CGGGCACAAATTCCATTCGCCCTTC............GACAAC......C 322
           |     ::::::    :::    |||:::                |||||| 
      759 sLysAsnGlnAlaTyrArgProLeuLeuLeuThrThrAspAsnGlyIleL 776

      323 ATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTC 372
             |||    |||||||||::::::    ||||||
      776 ysAlaTyrHisSerAspGlnGluAlaAlaGly.................. 786

      373 AGCCTTTACCGCATCCATTGGGACGGATACGAA......CACCATCCCGC 416
             |||    |||          |||       |||             ||
      787 ...LeuValArgTyrThrAsnAspArgGlyGluLeuIlePheThrAlaAl 802

      417 CGAC......GGCTATGACGGGCCACAGGGCGGCGGCTATCCC....... 453
           ||||        ||||||    :::||||||    :::||||||
      802 aAspIleLysGlyTyrAlaAsnProGlnValSerGlyTyrLeuGlyValT 819

      454 ..GCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTT 501
             :::|||    ||||||                              ::::::
      819 rpValProValGlyAla..........................AlaAla 826

      502 GCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACG 551
             |||::::::|||:::          :::          ||||||       ::
      827 AspGlnAspValArgValAlaAlaSerThrAlaProSerThrAspGlyLy 843

      552 GCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAG 601
           :                |||:::    :::|||:::    ::::::    :::
      843 s.........SerValHisGlnAsnAlaAlaLeuAspSerArgValMetP 857

      602 GCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCG 651
```

```
       :::|||||||::::::           ::::::   :::   ::::::
 857 heGluGlyPheSerAsnPheGlnAlaPheAlaThrLysLysGluGluTyr 873

 652 GGCAATGCCGCCGAAGCCTTCAAC.....................GGCAC 680
        |||::::::  |||   |||                   |||::
 874 ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluTrpGlyVa 890

 681 TGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCG 730
     :::::|||                :::      |||   :::   |||:::::
 890 lThrAsp...............PheGluMetAlaProGlnTyrValSerS 902

 731 CAGGCGATGCCGTG.........CAGGGTATAAGCGAAGGCTCAAACATT 771
     ::    |||:::            :::   |||::::::|||        :::
 902 erThrAspGlySerPheLeuAspSerValIleGlnAsnGlyTyrAlaPhe 918

 772 GCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCG 821
     :::      :::   ||||||    :::|||       ||||||   :::
 919 ThrAspArgTyrAspLeuGly...IleSerLysProAsnLysTyrGlyTh 934

 822 CATCAACGATTTGGCAGAT...ATGGCGCAACTCAAAGACTATGCCGCAG 868
          :::||||||:::      :::       |||   :::   :::
 934 rAlaAspAspLeuValLysAlaIleLysAlaLeuHisSerLysGlyIleL 951

 869 CAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATA 918
       ::::::   |||||||:::   ::::::
 951 ysValMetAlaAspTrpValProAspGln..................... 960

 919 GAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGG...AT 965
                               |||   |||::::|||   |||      ::
 961 ....................MetTyrAlaLeuProGluLysGluValVa 970

 966 TGGAGCTGTCCGG...GGAAAATACGGCTTGGGCGGCATCACGGCACATC 1012
      :     |||::::|||      |||||||||                    |
 970 lThrAlaThrArgValAspLysTyrGlyThr..................P 981

1013 CTGTCAAGCGGTCGCAGATG...GGCGCGATCGCATTGCCGAAAGGGAAA 1059
      |||||     ||||||:::   :::::::::   :::       ||||||
 981 roValAlaGlySerGlnIleLysAsnThrLeuTyrValValAspGlyLys 997

1060 TCCGCCGTCAGCGACAAT.....................TTTGCCGATGC 1088
      |||:::    :::|||:::                    |||     :::
 998 SerSerGlyLysAspGlnGlnAlaLysTyrGlyGlyAlaPheLeuGluGl 1014

1089 GGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAA 1138
               ||||||||||||:::      :::      ::::::|||
1014 uLeuGlnAlaLysTyrProGluLeuPheAlaArgLysGlnIle....... 1028

1139 ACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCG 1188
                                              |||:::    ||||||
1029 ..................................SerThrGlyValPro 1033

1189 ......CCGTCA...............................AACGG 1199
            ||||||                                 |||||
1034 MetAspProSerValLysIleLysGlnTrpSerAlaLysTyrPheAsnGl 1050

1200 CAAAAATGTC....................AAACTGGCAGACCAACGCC 1228
      |    |||:::                    |||     ||||||
1050 yThrAsnIleLeuGlyArgGlyAlaGlyTyrValLeuLysAspGlnAlaT 1067

1229 AC...................................CCGAAG... 1236
                                            ||||||
1067 hrAsnThrTyrPheSerLeuValSerAspAsnThrPheLeuProLysSer 1083

1237 ..................................ACAGGCGTACCGTT 1250
                                       |||||||:::    ||
1084 LeuValAsnProAsnHisGlyThrSerSerSerValThrGlyLeuValPh 1100

1251 TGACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACG 1299
      ||||||||||||||:::                   |||    |||:::|||
1100 eAspGlyLysGlyTyr..................ValTyrTyrSerThr 1110

seq_name: SwissProt_40:SMF1_HUMAN
```

```
seq_documentation_block:
ID   SMF1_HUMAN     STANDARD;      PRT;  1902 AA.
AC   O14497; Q9UPZ1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   SWI/SNF-related, matrix-associated, actin-dependent regulator of
DE   chromatin subfamily F member 1 (SWI-SNF complex protein p270) (B120).
GN   SMARCF1 OR C1ORF4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS.
RX   MEDLINE=20221560; PubMed=10757798;
RA   Dallas P.B., Pacchione S., Wilsker D., Bowrin V., Kobayashi R.,
RA   Moran E.;
RT   "The human SWI-SNF complex protein p270 is an ARID family member with
RT   non-sequence-specific DNA binding activity.";
RL   Mol. Cell. Biol. 20:3137-3146(2000).
RN   [2]
RP   SEQUENCE OF 1-1175 FROM N.A.
RX   MEDLINE=98094256; PubMed=9434167;
RA   Takeuchi T., Chen B.-K., Qiu Y., Sonobe H., Ohtsuki Y.;
RT   "Molecular cloning and expression of a novel human cDNA containing CAG
RT   repeats.";
RL   Gene 204:71-77(1997).
RN   [3]
RP   SEQUENCE OF 1-1132 FROM N.A.
RA   Takeuchi T., Misaki A.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS DNA NON-SPECIFICALLY.
CC   -!- SUBUNIT: PART OF THE SWI-SNF COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN SPLEEN, THYMUS, PROSTATE,
CC       TESTIS, OVARY, SMALL INTESTINE, COLON, AND PBL, AND AT A MUCH
CC       LOWER LEVEL IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
CC       MUSCLE, KIDNEY, AND PANCREAS.
CC   -!- SIMILARITY: CONTAINS 1 ARID DOMAIN.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1132
CC       ONWARD DUE TO A FRAMESHIFT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF265208; AAF75765.1; ALT_INIT.
DR   EMBL; AB001895; BAA23269.1; ALT_FRAME.
DR   EMBL; AB024075; BAA83073.1; ALT_SEQ.
DR   EMBL; AB024059; BAA83073.1; JOINED.
DR   EMBL; AB024060; BAA83073.1; JOINED.
DR   EMBL; AB024061; BAA83073.1; JOINED.
DR   EMBL; AB024062; BAA83073.1; JOINED.
DR   EMBL; AB024063; BAA83073.1; JOINED.
DR   EMBL; AB024064; BAA83073.1; JOINED.
DR   EMBL; AB024065; BAA83073.1; JOINED.
DR   EMBL; AB024066; BAA83073.1; JOINED.
DR   EMBL; AB024067; BAA83073.1; JOINED.
DR   EMBL; AB024069; BAA83073.1; JOINED.
DR   EMBL; AB024068; BAA83073.1; JOINED.
DR   EMBL; AB024070; BAA83073.1; JOINED.
DR   EMBL; AB024071; BAA83073.1; JOINED.
DR   EMBL; AB024072; BAA83073.1; JOINED.
DR   EMBL; AB024073; BAA83073.1; JOINED.
DR   EMBL; AB024074; BAA83073.1; JOINED.
DR   MIM; 603024; -.
DR   InterPro; IPR001606; ARID.
DR   Pfam; PF01388; ARID; 1.
```

```
DR   SMART; SM00501; BRIGHT; 1.
KW   DNA-binding; Nuclear protein.
FT   DOMAIN       96     99       POLY-GLN.
FT   DOMAIN      178    184       POLY-GLN.
FT   DOMAIN      615    618       POLY-SER.
FT   DOMAIN      631    740       ARID.
FT   DOMAIN      944   1021       GLN-RICH.
FT   MUTAGEN     690    690       W->A: PARTIAL LOSS OF DNA-BINDING
FT                                ACTIVITY. WHEN A.A. 713 ALSO MUTATED,
FT                                COMPLETE LOSS OF ACTIVITY.
FT   MUTAGEN     713    713       Y->A: PARTIAL LOSS OF DNA-BINDING
FT                                ACTIVITY. WHEN A.A. 690 ALSO MUTATED,
FT                                COMPLETE LOSS OF ACTIVITY.
FT   CONFLICT     27     27       D -> G (IN REF. 1).
FT   CONFLICT     51     51       V -> M (IN REF. 1).
FT   CONFLICT    349    349       Q -> S (IN REF. 2).
FT   CONFLICT    368    368       G -> GG (IN REF. 3).
FT   CONFLICT    374    374       P -> S (IN REF. 2).
FT   CONFLICT    393    393       P -> L (IN REF. 2).
FT   CONFLICT    489    502       MANMPPQVGSGMCP -> NGQYATSGWVRDVS (IN REF.
FT                                2).
FT   CONFLICT    634    634       E -> G (IN REF. 2).
FT   CONFLICT    924    924       P -> S (IN REF. 2).
FT   CONFLICT   1016   1016       Q -> L (IN REF. 2).
FT   CONFLICT   1033   1033       Q -> P (IN REF. 2).
SQ   SEQUENCE   1902 AA;  205946 MW;  9B70A46F81062EAA CRC64;

alignment_scores:
           Quality:  123.00          Length:     502
             Ratio:  0.594             Gaps:      25
Percent Similarity:  41.235   Percent Identity: 21.912

alignment_block:
 US-09-303-518D-463 x SMF1_HUMAN  ..

 Align seg 1/1  to: SMF1_HUMAN  from: 1  to: 1902

   24 CCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATGCATGCACACGCCT 73
       |||||||||        :::|||            :::     :::::
  876 ProTyrSerArgAlaAlaGly............ProGlyLeuGlyAsnVa 888

   74 CAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCAT 123
       :   :::|||   |||:::    ||||||   |||:::     :::
  888 lAlaMetGlyProArgGlnHisTyrProTyrGlyGlyProTyrAspArgV 905

  124 TTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGA...GCT 170
         ||||||:::            |||   |||     :::|||:::   :::
  905 alArgThrGlu.........ProGlyIleGlyProGluGlyAsnMetSer 918

  171 TGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACA........... 209
            |||   :::|||        :::   ::::::
  919 ThrGlyAlaProGlnProAsnLeuMetProSerAsnProAspSerGlyMe 935

  210 ....AAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAA 255
          :::||||||       ||||||::::: ||||||        :::
  935 tTyrSerProSerArgTyrProProGlnGln.GlnGlnGlnGlnGlnGln 951

  256 GGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCA 305
         :::   :::|||      :::||||||    :::|||
  952 ArgHisAspSerTyrGlyAsnGlnPheSerThrGlnGlyThrProSerGl 968

  306 TTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTA 355
        |||||||||
  968 ySerProPhe........................................ 971

  356 GTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAA 405
        |||   :::      ::::::|||:::    :::    ::::::|||:::
  972 ..ProSerGlnGlnThrThrMetTyrGlnGlnGlnGlnGlnAsnTyrLys 987

  406 CACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGC 455
         :::|||   ||||||      ||||||               |||||
  988 ...ArgProMetAspGlyThrTyrGlyPro...............ProAl 998

  456 TCCCAAAGGCGCGAGGGATATATACAGCTACGA................. 488
      |    :::           ::::::||||||
  998 aLysArgHisGluGlyGluMetTyrSerValProTyrSerThrGlyGlnG 1015

  489 ........CATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGA 530
               ::::::::::::   |||          |||||||||
 1015 lyGlnProGlnGlnGlnGlnLeuProProAlaGlnProGlnProAlaSer 1031

  531 CAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCG 580
       |||   |||            |||::::::|||    ||||||
 1032 GlnGlnGln............AlaAlaGlnProSerProGlnGlnAspVa 1044

  581 CTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATAC 630
       |||                ::::::    :::||||||   |||   ||
 1044 lTyr...............AsnGlnTyr..GlyAsnAlaTyrProAlaTh 1055

  631 AGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCAC 680
      ||||    :::    |||:::|||                |||::::::   :
 1055 rAlaThrAlaAlaThrGluArgArgProAlaGlyGlyProGlnAsnGlnP 1072

  681 TG.....................CAGATATCGTCAAAAACATCATCGGCG 709
      ::                      ::::::|||:::      ::::::
 1072 heProPheGlnPheGlyArgAspArgValSerAlaProProGlyThrAsn 1088

  710 CGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAA 759
         |||:::::::::        |||   |||          :::   |||::
 1089 AlaGlnGlnAsnMetProProGlnMetMetGlyGlyProIleGlnAlaSe 1105

  760 GGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAA 809
       :|||::::::
 1105 rAlaGluValAla..................................... 1109

  810 CAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACT 859
               :::     :::|||:::||||||         ||||||:::    |||
 1110 ........GlnGlnGlyThrMetTrpGln...GlyArgAsnAspMetThr 1122

  860 ..................................ATGCCGCAGCAGCCATC 876
                                        ||||||   |||::
 1123 TyrAsnTyrAlaAsnArgGlnSerThrGlySerAlaProGlnGlyProAl 1139

  877 CGCGATTGGGCAGTCCAAAACC...CCAATGCCGCACAAGGCATAGAAGC 923
      :          |||   ::::::|||         |||   |||          :::
 1139 aTyrHisGlyValAsnArgThrAspGluMetLeuHisThrAspGlnArgA 1156

  924 CGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTG 973
        :::       :::   |||
 1156 laAsnHisGluGlySerTrpPro........................... 1163

  974 TCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGG 1023

 1163 .................................................. 1163

 1024 TCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCG......CCGT 1067
                       ||||||      :::        |||        ||||
 1164 ...............SerHisGlyThrArgGlnProProTyrGlyProS 1175

 1068 CAGCGACAATTTTGCCGATG.............................. 1087
      |||||     :::    ||||||
 1175 erAlaProValProProMetThrArgProProProSerAsnTyrGlnPro 1191

 1088 ...............CGGCATACGCCAAATACC................. 1105
                      :::|||    |||::::::
 1192 ProProSerMetGlnAsnHisIleProGlnValSerSerProAlaProLe 1208

 1106 ....CGTCCCCTT.........ACCATTCCCGAAATATCCGTTCAAACTT 1142
           ||||||:::          |||   |||:::     |||
 1208 uProArgProMetGluAsnArgThrSerProSerLysSerProPheLeuH 1225

 1143 GGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGC... 1189
        |||::::::          :::|||::::::||||||    |||:::
 1225 isSerGlyMetLysMetGlnLysAlaGlyProProValProAlaSerHis 1241
```

```
1190 ................................CGTCAAACGGCAAAAAT 1206
                                     |||:::       ::
1242 IleAlaProAlaProValGlnProProMetIleArgArgAspIleThrPh 1258

1207 GTCAAACTGGCAGACCAACGCCACCCGA.........AGACAGGCGTACC 1247
      :          :::::::::||||||:::         :::|||     |
1258 eProProGlySerValGluAlaThrGlnProValLeuLysGlnArgArgA 1275

1248 GTTTGACGGTAAAG..................GGTTTCCGAATTTTGAGA 1279
     ||||||||:::|||                   :::::|||::::::
1275 rgLeuThrMetLysAspIleGlyThrProGluAlaTrpArgValMetMet 1291

1280 AGC 1282
     |||
1292 Ser 1292

seq_name: SwissProt_40:V70K_TYMVA

seq_documentation_block:
ID   V70K_TYMVA     STANDARD;      PRT;   628 AA.
AC   P20131;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   69 kDa protein.
OS   Turnip yellow mosaic virus (Australian isolate).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX   NCBI_TaxID=12155;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90021184; PubMed=2800335;
RA   Keese P., Mackenzie A., Gibbs A.;
RT   "Nucleotide sequence of the genome of an Australian isolate of turnip
RT   yellow mosaic tymovirus.";
RL   Virology 172:536-546(1989).
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J04373; AAA46591.1; -.
DR   PIR; JQ0110; JQ0110.
SQ   SEQUENCE   628 AA;  68740 MW;  67CD342A09161D96 CRC64;

alignment_scores:
           Quality:  122.50           Length:     445
             Ratio:  0.613              Gaps:      31
  Percent Similarity:  44.944   Percent Identity: 26.517

alignment_block:
 US-09-303-518D-463 x V70K_TYMVA   ..

 Align seg 1/1  to: V70K_TYMVA  from: 1  to: 628

      8 TTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCG 57
         |||         :::||||||     |||      :::       ||
     78 PheSerProLeuHisProLeuValPheProGlyHisHisSerGlnLeuAr 94

     58 ATGCATGCAC...ACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCA 104
         |    :::|||    |||   |||:::   ||||||   |||
     94 gHisValHisGluThrGlnGlnValGlnGlnThrCysPro.......... 107

    105 GGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATT.. 152
           :::    :::::::::|||:::     :::|||:::||||||
    108 .....GlyGluLeuLysLeuSerGly.IleGluGluLeuProProAlaPr 122

    153 ....CGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTG 198
          ||||||::::::                      ||||||   |||:::
    122 oGlnArgGlnHisSerLeuProLeuHisIleThrArgProSerArgPheP 139

    199 GGAAACAT.........ACAAAGCCATCAGTTGGGCCA......CCTGAT 233
          |||                :::|||:::|||   |||      ||||||
    139 roHisHisPheHisAlaArgArgProAspValLeuProSerLeuProAsp 155

    234 GATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTT 282
                                              |||   |||:::
    156 ..................................HisGlyProValLe 160

    283 ............CCGATCACGGGCACAAATTCCATTCGCCCTTCGACAA 320
                    |||   |||::::::::::::    |||    :::|||
    160 uAlaGluThrLysProArgThrSerValArgGlnProArgSerThrThrA 177

    321 CCATGCCTCACATT...CCGATTCTGACGAAGCCGGTAGTCCCGTTGACG 367
          |||    :::    |||||||||             ||||||    :::
    177 rgGlyProSerPheArgProIleLeuLeuProGluValValHisValHis 193

    368 GATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCC 417
        |||  :::|||    ||||||                         |||
    194 Asp.AspPro...ProHis.......................SerSerL 201

    418 GACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGC 467
          |||        |||:::    :::|||:::|||:::    :::    ||||||
    201 euArgProGlyArgSerArgSerArgGlnLeuGlnProThrIleArgArg 217

    468 GAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCC 517
                 :::          ::::::    :::    |||:::    |||||
    218 Pro......LeuLeuAlaProAsnGlnPheAsnSerProArgGlnProPr 232

    518 TCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTC 567
        |    |||:::|||:::
    232 oPro.LeuSerAspAsp................................ 237

    568 CACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACG 617
                                              |||:::      :
    238 ..................................ProGlyIleLeu.G 242

    618 CGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAG 667
        :::|||          ||||||:::    |||    |||      ||||||:::
    242 lyProCysProLeuAlaProAsnSerThr...ArgAspProProProArg 257

    668 CCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGA 717
        |||    |||              ::::::        |||    ::::::|||
    258 ProIleThrProGlyProPheAsn......ThrHisGlyValArgProLe 272

    718 GAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAA 767
          ::::::          :::|||    |||    |||
    272 uSerValLeuProArgAlaProProArgArg.................. 282

    768 CATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGG 817
                                          :::|||          |||
    283 ......................GlyLeuLeuProAsnProArgArgHis 291

    818 CGCGCATCAACG......ATTTGGCAGATATGGCGCAACTCAAAGACTAT 861
        |||:::||||||      |||    :::            :::      |||
    292 ArgThrSerThrGlyHisIleProSerThrThrAlaSerArgProThrGl 308

    862 GCCG...........CAGCAGCCATCCGCGATTGGGCAGTCCAAAACCC 899
          |||               |||:::|||          :::    ||||||
    308 yProProSerArgLeuGlnArgProValHisLeuHisGlnSerGlyProH 325

    900 CAATGCCGCACAAGGCATAGAAGCCGTCAGCAATA.........TCTTTA 940
              |||               :::|||||||:::|||         ::::::
    325 isThrProAspPhe......ArgProSerGlyIleArgLysAspAlaPhe 339

    941 TGGCAGCCATCCCCATCAAAGGGA...TTGGAGCTGTCCGGGGAAAATAC 987
           |||    :::|||         |||    ||||||      |||:::::::::
    340 LeuGlnThrGlyProArgLeuGlyHisLeuGluArgLeuGlyGlnSerAl 356

    988 GGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGC 1037
```

```
      :        :::|||:::::     |||:::   ||||||       |
 356 aAspLeuArgThrSerGluArgThrProSerThrLysArgArgLeuProA 373

1038 GA.........TCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATT 1078
     ||         :::|||       ::::::   |||:::||||||:::
 373 rgProSerGluProAsnCysLeuProSerSerLeuProGluAlaThrLeu 389

1079 TTGCCG...ATGCGGCAT.................ACGCCAAATACCCG 1107
        |||      ||||||                    ||||||
 390 AlaProSerTyrArgHisArgArgSerHisProLeuLeuProAsnPro.. 405

1108 TCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGG 1157
       |||   :::::|||   |||::: :::        |||        |||
 406 .ProAlaAlaLeuProProIleAla.TyrThrSerGly.........Arg 418

1158 CAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATG 1207
        :::    ||||||   |||    ::::::|||:::    ::::::
 419 GlyLysIleHisHisSerLeuProLysGlyAlaLeuProLysGlu..... 433

1208 TCAAACTGGCAGACCAACGCCACCC 1232
            |||    |||   ||||||
 434 .......GlyProProProProPro 439
```

seq_name: SwissProt_40:NCR1_HUMAN

```
seq_documentation_block:
ID   NCR1_HUMAN     STANDARD;      PRT;  2440 AA.
AC   O75376; Q9UPV5; Q9UQ18;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear receptor co-repressor 1 (N-CoR1) (N-CoR).
GN   NCOR1 OR KIAA1047.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98393736; PubMed=9724795;
RA   Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;
RT   "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
RT   transcription by interaction with the human N-CoR/mSin3/HDAC1
RT   complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
RN   [2]
RP   SEQUENCE OF 782-2440 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K.-I., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
RN   [3]
RP   SEQUENCE OF 974-2440 FROM N.A.
RX   MEDLINE=99375328; PubMed=10444336;
RA   Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,
RA   Horwitz K.B., Lupski J.R., Seo H.;
RT   "Localization of the human nuclear receptor co-repressor (hN-CoR) gene
RT   between the CMT1A and the SMS critical regions of chromosome
RT   17p11.2.";
RL   Genomics 59:339-341(1999).
CC   -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC       NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC       PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC   -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC       AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC       WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
CC       ABSENCE OF LIGAND.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC       ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC       AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC       INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC       DOMAINS (ID1 AND ID2).
CC   -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC       SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
CC       SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC       FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC       SPECIFICITY.
CC   -!- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 CORNR BOX.
CC   -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF044209; AAC33550.1; -.
DR   EMBL; AB028970; BAA82999.1; -.
DR   EMBL; AB019524; BAA75814.1; -.
DR   MIM; 600849; -.
DR   InterPro; IPR001005; Myb_DNA_bind.
DR   Pfam; PF00249; myb_DNA-binding; 2.
DR   SMART; SM00395; SANT; 2.
DR   PROSITE; PS50090; MYB_3; 1.
KW   Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW   Coiled coil.
FT   DOMAIN      174    216       COILED COIL (POTENTIAL).
FT   DOMAIN      254    312       INTERACTION WITH SIN3A/B.
FT   DOMAIN      299    328       COILED COIL (POTENTIAL).
FT   DNA_BIND    437    482       SANT-A (POTENTIAL).
FT   DNA_BIND    620    670       MYB.
FT   DOMAIN      501    557       COILED COIL (POTENTIAL).
FT   DOMAIN      607    617       PRO-RICH.
FT   DOMAIN      988   1816       INTERACTION WITH ETO.
FT   DOMAIN     2055   2059       CORNR BOX OF ID1.
FT   DOMAIN     2263   2267       CORNR BOX OF ID2.
FT   DOMAIN       58     64       POLY-GLN.
FT   DOMAIN      593    603       POLY-ALA.
FT   DOMAIN     1032   1035       POLY-PRO.
FT   DOMAIN     1707   1712       POLY-ALA.
FT   DOMAIN     1952   1963       POLY-SER.
FT   CONFLICT   1014   1014       L -> V (IN REF. 2).
FT   CONFLICT   1508   1509       PP -> SS (IN REF. 2).
FT   CONFLICT   1561   1561       W -> R (IN REF. 2).
FT   CONFLICT   1567   1567       Q -> H (IN REF. 2).
SQ   SEQUENCE   2440 AA;  270263 MW;  60A4D7964D00EDAB CRC64;
```

```
alignment_scores:
           Quality:  121.50          Length:    646
             Ratio:  0.428             Gaps:     31
 Percent Similarity:  43.963   Percent Identity: 19.969

alignment_block:
 US-09-303-518D-463 x NCR1_HUMAN   ..

 Align seg 1/1  to: NCR1_HUMAN  from: 1  to: 2440

  12 CCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTG......CCTGC 55
       |||    :::::|||          |||    ::::::|||      |||
 924 ProLeuAspLeuProGlnLeuGlnHisArgAlaAlaValIleProProMe 940

  56 CGATGCATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAG 105
         :::||||||         ::::::                 |||  |
 940 tValSerCysThrProCysAsnIle.................ProIleG 951
```

```
 106 GTTCTCGACCGTCAGCATTT....................CGAACCCGA 134
     ||:::       |||:::::                    ::::::
 951 lyThrProValSerGlyTyrAlaLeuTyrGlnArgHisIleLysAlaMet 967

 135 CGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACG 184
     :::|||          :::::::::|||                |||||
 968 HisGluSerAlaLeuLeuGluGluGln...............ArgGlnAr 979

 185 GCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATG 234
     |        :::|||       :::     :::    |||:::::::|||
 979 gGlnGluGlnIleAspLeuGluCysArgSerSerThrSerProCysGlyT 996

 235 ATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTC 284
      |||   :::    |||    :::|||   ::::::          |||
 996 hrSer..LysSerProAsnArgGluTrpGluValLeuGlnProAlaProH 1012

 285 CGATCACGGGCACAAAT.........TCCATTCGCCCTTCGACAACCATG 325
                 ||||||        :::::::|||     ||||||
1012 isGlnLeuIleThrAsnLeuProGluGlyValArgLeuProThrThrArg 1028

 326 CCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGC 375
     |||              |||:::|||      ||||||      |||::
1029 Pro..............ThrArgProProProProLeuIleProSerSe 1040

 376 CTTTACCGCATCCATT..........................GGGACGG 398
     :   |||:::                                |||:::
1040 rLysThrThrValAlaSerGluLysProSerPheIleMetGlyGlySerI 1057

 399 ATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCT 448
        ::::::        |||   |||    :::|||:::|||:::   |||:::
1057 leSerGlnGlyThrProGlyThrTyrLeuThrSerHisAsnGlnAlaSer 1073

 449 ATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGC 498
                                 ||||||     |||
1074 ..........................TyrThrGlnGluThr........ 1078

 499 GTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACA 548
         |||||                      |||:::::::::::    :::
1079 ....ProLys.................ProSerValGlySerIleSerL 1088

 549 ACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAG 598
       |||||||||                              ::::::|||
1088 euGlyLeuProArg.........................GlnGlnGlu 1095

 599 TAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGA 648
       |||   :::::::::    |||                :::::::::
1096 SerAlaLysSerAlaThrLeuProTyrIleLysGlnGluGluPheSerPr 1112

 649 TCGGGCAATGCCGCCGAAGCCT........................... 670
      |||::::::         :::|||
1112 oArgSerGlnAsnSerGlnProGluGlyLeuLeuValArgAlaGlnHisG 1129

 671 ....................TCAACGGCACTGCAGATATCGTCAAAAACA 700
                           :::    |||:::|||    :::|||    |||
1129 luGlyValValArgGlyThrAlaGlyAlaIleGlnGluGlySerIleThr 1145

 701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
         ::::::        :::|||:::|||::::::::::::|||
1146 ArgGlyThrProThrSerLysIleSerValGluSerIlePro........ 1159

 751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGTCTGCTTTC 800

1159 .................................................. 1159

 801 CACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 850
                                 |||    :::|||    ::::::
1160 ......................SerLeuArgGlySerIleThrGlnGly 1168

 851 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900
     :::    ::::::||||||                           |||
1169 ThrProAlaLeuProGln..........................ThrGl 1176

 901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT.........TTAT 941
      :::|||   :::|||          ::::::||||||           :::
1176 yIleProThrGluAlaLeuValLysGlySerIleSerArgMetProIleG 1193

 942 GGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCT 991
      :::     ||||||:::||||||    |||     :::
1193 luAspSerSerProGluLysGlyArgGluGluAla.............. 1204

 992 TGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATC 1041
        |||::::::    |||::::::    :::|||:::
1205 ...AlaSerLysGlyHisValIleTyrGluGlyLys............. 1215

1042 GCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGC 1091
                                |||:::    ||||||         :
1216 ........................SerGlyHisIleLeuSerTyrAspA 1224

1092 ATACGCCAAATACCCGT...............CCCCTTACCATTCCCGAA 1126
     ::        |||:::|||                  |||   |||      |||
1224 snIleLysAsnAlaArgGluGlyThrArgSerProArgThrAlaHisGlu 1240

1127 ATATCCGTT...CAAACTTGGAGCAGCGTTACGGC.......AAAGAAAA 1166
     ||||||:::    :::::::::::||||||              |||:::::
1241 IleSerLeuLysArgSerTyrGluSerValGluGlyAsnIleLysGlnGl 1257

1167 CATCACCTCCTCAACCGTGCCG.........CCGTCAAACGGCAAAAATG 1207
     :::::::                  |||         |||    :::|||
1257 yMetSerMetArgGluSerProValSerAlaProLeuGluGlyLeuIleC 1274

1208 TCAAACTGGCAGACCAACGCCACCCG............AAGACAGGCGTA 1245
       :::        :::      |||               |||      :::
1274 ysArgAlaLeuProArgGlySerProHisSerAspLeuLysGluArgThr 1290

1246 CCGTTTGACGGT.........AAAGGGTTTCCG...............AA 1271
        ::::::|||          :::|||    |||                 ::
1291 ValLeuSerGlySerIleMetGlnGlyThrProArgAlaThrThrGluSe 1307

1272 TTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAA....... 1314
     :||||||        :::||||||      ::::::       :::
1307 rPheGluAspGlyLeuLysTyrProLysGlnIleLysArgGluSerProP 1324

1315 .....GAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGAT 1359
           :::          |||:::|||   |||:::||||||
1324 roIleArgAlaPheGluGlyAlaIleThrLysGlyLysProTyrAspGly 1340

1360 GCGAAACCGAGATGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCG 1409
                    |||:::    |||:::::::::::::::        ::
1341 IleThrThrIleLysGluMetGlyArgSerIleHisGluIleProArgGl 1357

1410 TGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTA 1459
     ::::   :::    ::::::    ::::::|||    :::        ||||||:
1357 nAspIleLeuThrGlnGluSerArgLysThrProGluValValGlnSerT 1374

1460 GTCAG.................TTTAAAGCCCATGCGCAACGAGAATGG 1491
     :::::                  ::::::       :::       ::::::
1374 hrArgProIleIleGluGlySerIleSerGlnGlyThrProIleLysPhe 1390

1492 GAAAATAAAACAGGGTTAGAT...TTTAATCATTTTATAGGTGGTGATAT 1538
     :::|||::::::|||      :::    ::::::|||
1391 AspAsnAsnSerGlyGlnSerAlaIleLysHis................ 1401

1539 CAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGT.... 1584
       |||    |||:::    :::||||||        :::|||:::||||||
1402 .AsnValLysSerLeuIleThrGlyProSerLysLeuSerArgGlyMetP 1418

1585 ..GATGTACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTT 1632
         ::::::::::::    ::::::              ::::::|||
1418 roProLeuGluIleValProGluAsnIleLysValValGluArgGlyLys 1434

1633 TATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTGAA 1682
     |||:::
1435 TyrGlu.......................................As 1437

1683 AACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACC 1719
```

```
                :::|||  |||     |||  ::::::||||||
      1437 pValLysAlaGlyGluThrValArgSerArgHisThr 1449

seq_name: SwissProt_40:PSC_DROME

seq_documentation_block:
ID   PSC_DROME      STANDARD;      PRT;  1603 AA.
AC   P35820;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Posterior sex combs protein.
GN   PSC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92018190; PubMed=1833647;
RA   Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT   "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RT   encode proteins with homology to the murine bmi-1 oncogene.";
RL   Nature 353:351-353(1991).
CC   -!- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC       EXPRESSION PATTERNS OF THE HOMEOTIC SELECTOR GENES OF THE
CC       ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
CC       THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X59275; CAA41965.1; -.
DR   PIR; S17983; S17983.
DR   FlyBase; FBgn0005624; Psc.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT   DOMAIN       47     53       POLY-THR.
FT   DOMAIN       83     88       POLY-THR.
FT   DOMAIN       91     98       POLY-THR.
FT   DOMAIN      145    152       POLY-THR.
FT   DOMAIN      184    202       POLY-SER.
FT   ZN_FING     265    304       RING-TYPE.
FT   DOMAIN      642    651       POLY-SER.
FT   DOMAIN     1066   1069       POLY-GLY.
FT   DOMAIN     1185   1189       POLY-PRO.
FT   DOMAIN     1214   1217       POLY-PRO.
FT   DOMAIN     1391   1396       POLY-PRO.
FT   DOMAIN     1458   1461       POLY-ALA.
FT   DOMAIN     1517   1520       POLY-GLY.
SQ   SEQUENCE   1603 AA;  169999 MW;  77024F4097736473 CRC64;

alignment_scores:
          Quality:  121.00           Length:    361
            Ratio:  0.761              Gaps:     15
 Percent Similarity:  44.044   Percent Identity: 22.161

alignment_block:
 US-09-303-518D-463 x PSC_DROME

Align seg 1/1  to: PSC_DROME  from: 1  to: 1603

 275 TCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCAT 324
     |||:::  |||   :::|||   :::     :::   |||   :::
1128 SerSerSerProAsnAsnGlyGlnHisProLysHisLysSerProValAs 1144

 325 GCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAG 374
        :::|||  |||:::       |||     |||      ::::::
1144 nAsnTyrIleGluIleValLysLeuProAspGlnPro......GlnAspG 1159

 375 CCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCT 424
           |||:::     ::::::  :::      |||||||:::|||
1159 lnValGlnAlaAlaLysGluAlaGlnLysArgGlnSerProProAlaAla 1175

 425 ATGACGGGCCACAGGGCGGCGGCTATCCCG.................... 454
     :::   ||||||  ||||||  :::|||
1176 ValProGlyHisLeuAlaAlaLysLeuProProProProProSerLysAl 1192

 455 .CTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGC 503
      :::||||:::   :::   :::   :::          |||      ||||
1192 aIleProSerProGlnHisLeuValSerArgMetThrProProGlnLeuP 1209

 504 CCAAAATATCCGCCTCA..........................ACCTGA 526
     ||||||:::::::::                            |||
1209 roLysValAlaThrProProProProSerSerProArgValIleThrPro 1225

 527 CCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCC 576
     |||  |||:::  |||  |||:::      ::::::          ||
1226 ProLysThrSerProProAlaAsnAlaAlaLysValThrProLeuLysPr 1242

 577 GGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCG 626
     |:::|||
1242 oValLeu........................................... 1244

 627 ATACAGCCCCGAGCTGGACAGATCGGGCAATGCCG...CCGAAGCCTTCA 673
          :::|||:::  :::|||:::     |||   |||:::
1245 .....ThrProThrGlnValAspLysLysThrProSerProGluLysArg 1259

 674 ACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATT 722
     ||||||  |||::::::|||   :::   :::|||   ::::::|||
1260 ThrAlaAlaGlnMetGlySerHisSerProThrAlaSerGluAsnLysSe 1276

 723 ...............GTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCG 757
                  :::|||::::::::::::::::  |||   ::::
1276 rProLysGlyGlyAlaAlaGlyValAlaAsnSerThrGlyGlyThrGlnA 1293

 758 AAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAA 807
     :::|||:::   |||                           :::   :::
1293 snGlyAspProAlaAla....................LysLysPheArg 1302

 808 AACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGA 857
          :::   |||   |||   :::   ::::::|||
1303 ProIleLeuProArgGlnAsnGlyMetProGluLeuAlaProLysLeuPr 1319

 858 CTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCG 907
         |||        :::          :::|||||||||   |||:
1319 oThrLeuAlaProPheValGlyPheAsnProLeuGlnAsnProAlaAlaG 1336

 908 CACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATC 957
     :::::  :::     ||||||               :::   ||||||::
1336 lyLysLysVal.ProProSer..............LysLysSerProAs 1347

 958 AAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGC 1007
     :   |||                                  |||:::    |
1347 nAlaGly..............................AlaAlaAlaH 1353

1008 ACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGA 1057
     ||      :::|||||||:::          ::::::      :::   |||
1353 is......GlnSerGlyGln.........GlnLysLeuValAsnGlyGly 1364

1058 AATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCG 1107
     :::|||    ||||||
1365 GlnProGlnSerAla.................................. 1369
```

```
  1108 TCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGG 1157

  1370 ..............................................G 1370

  1158 CAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATG 1207
          :::||||||||||||||          :::::::|||   ::::|||
  1370 lnGlnLysThrSerProProAlaGlnLysAsnGlnGlnGlnValLysLys 1386

  1208 TCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGT 1257
           :::    :::||||||||| |||      :::::::|||      |||
  1387 .ValSerLysAsnProThrPro.ProProProSerLeuProAlaValGly 1402

  1258 AAAGGGTTTCCGAATTTTGAGAAGCAC 1284
        |||    :::|||:::          |||
  1403 LysMetMetProHisProValMetHis 1411
```

seq_name: SwissProt_40:N214_HUMAN

```
seq_documentation_block:
ID   N214_HUMAN     STANDARD;      PRT;  2090 AA.
AC   P35658;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear pore complex protein Nup214 (Nucleoporin Nup214) (214 kDa
DE   nucleoporin) (CAN protein).
GN   NUP214 OR CAN OR CAIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=92195315; PubMed=1549122;
RA   Von Lindern M., Fornerod M., Van Baal S., Jaegle M., De Wit T.,
RA   Buijs A., Grosveld G.;
RT   "The translocation (6;9), associated with a specific subtype of acute
RT   myeloid leukemia, results in the fusion of two genes, dek and can,
RT   and the expression of a chimeric, leukemia-specific dek-can mRNA.";
RL   Mol. Cell. Biol. 12:1687-1697(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=94151361; PubMed=8108440;
RA   Kraemer D., Wozniak R.W., Blobel G., Radu A.;
RT   "The human CAN protein, a putative oncogene product associated with
RT   myeloid leukemogenesis, is a nuclear pore complex protein that faces
RT   the cytoplasm.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:1519-1523(1994).
CC   -!- FUNCTION: MAY SERVE AS A DOCKING SITE IN THE RECEPTOR-MEDIATED
CC       IMPORT OF SUBSTRATES ACROSS THE NUCLEAR PORE COMPLEX.
CC   -!- SUBUNIT: HOMODIMER. INTERACTS WITH DDX19.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. CYTOPLASMIC FILAMENTS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, SPLEEN, BONE MARROW,
CC       KIDNEY, BRAIN AND TESTIS, BUT HARDLY IN ALL OTHER TISSUES OR IN
CC       WHOLE EMBRYOS DURING DEVELOPMENT.
CC   -!- DOMAIN: CONTAINS MANY X-X-F-G REPEATS.
CC   -!- PTM: PROBABLY GLYCOSYLATED AS IT REACT WITH WHEAT GERM AGGLUTININ
CC       (WGA).
CC   -!- DISEASE: IMPLICATED IN A SUBSET OF ACUTE MYELOID LEUKEMIA (ACUTE
CC       NONLYMPHOCYTIC LEUKEMIA) (AML) CARRYING A CHROMOSOMAL
CC       TRANSLOCATION T(6;9)(P23;Q34) THAT RESULTS IN THE FORMATION OF A
CC       DEK-CAN FUSION GENE.
CC   -!- DISEASE: IN A CASE OF ACUTE UNDIFFERENTIATED LEUKEMIA (AUL) A
CC       TRANSLOCATION RESULTS IN THE FORMATION OF A SET-CAN FUSION GENE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X64228; CAA45535.1; -.
DR   PIR; S26058; S26058.
DR   MIM; 114350; -.
DR   InterPro; IPR004325; Nucleoporin_FG.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF03093; Nucleoporin_FG; 17.
DR   SMART; SM00320; WD40; 1.
KW   Nuclear protein; Transport; Proto-oncogene; Chromosomal translocation;
KW   Repeat; Glycoprotein.
FT   DOMAIN      481   2076       11 X 5 AA APPROXIMATE REPEATS.
FT   DOMAIN     1409   2084       18 X 4 AA APPROXIMATE REPEATS.
FT   DOMAIN     1427   2085       11 X 3 AA APPROXIMATE REPEATS.
FT   DOMAIN     1213   2090       PRO/SER/THR-RICH.
FT   DOMAIN      740    768       LEUCINE-ZIPPER 1.
FT   DOMAIN      861    882       LEUCINE-ZIPPER 2.
FT   SITE        812    813       BREAKPOINT.
SQ   SEQUENCE   2090 AA;  213766 MW;  6DBE767FDD857F8F CRC64;
```

```
alignment_scores:
           Quality:  120.50             Length:    469
             Ratio:  0.606                Gaps:     19
 Percent Similarity:  42.431    Percent Identity: 21.109

alignment_block:
 US-09-303-518D-463 x N214_HUMAN   ..

 Align seg 1/1  to: N214_HUMAN  from: 1  to: 2090

    87 CGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACG 136
       |||   :::   |||:::|||:::      |||:::       ::::::
  1294 ArgProValAlaProSerGlyThrAlaLeuSerThrThrSerSerLysLe 1310

   137 GGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGC 186
        |||   |||                                       |
  1310 uGluThrPro.....................................P 1314

   187 CATATCGGATTGGGAAACATACAAAGCCATCA.................. 218
       ||    ::::::|||:::          ||||||
  1314 roSerLysLeuGlyGluLeuLeuPheProSerSerLeuAlaGlyGluThr 1330

   219 .......................GTTGGGCCACCTGATGATTCAACAGG 244
                              ||||||      |||||||||||
  1331 LeuGlySerPheSerGlyLeuArgValGlyGlnAlaAspAspSerThr.. 1346

   245 CGGCCGTTGAAGGAAATATCGGCTACA...TTGTCCGCTTTTCCGATCAC 291
       :::|||    ::::::    |||:::|||   |||::::::        ||
  1347 LysProThrAsnLysAlaSerSerThrSerLeuThrSerThrGlnProTh 1363

   292 GGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTC 341
       |   |||:::::::::      :::    ::::::          |||:::|
  1363 rLysThrSerGlyValProSerGlyPheAsnPheThrAlaProProValL 1380

   342 TG.........ACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACC 382
       ||             |||:::|||      |||:::|||:::||||||   |||
  1380 euGlyLysHisThrGluPro......ProValThrSerSerAlaThrThr 1394

   383 GCATCCATT......................................... 391
       :::|||:::
  1395 ThrSerValAlaProProAlaAlaThrSerThrSerSerThrAlaValPh 1411

   391 .................................................. 391

  1411 eGlySerLeuProValThrSerAlaGlySerSerGlyValIleSerPheG 1428

   392 ..GGGACGGAT....................................... 400
          ||||||:::
  1428 lyGlyThrSerLeuSerAlaGlyLysThrSerPheSerPheGlySerGln 1444

   401 ......ACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGC 444
             ::::::|||:::|||||||:::|||              :::||
```

```
1445 GlnThrAsnSerThrValProProSerAla...ProProProThrThrAl 1460

 445 GGCTATCCCGCTCCCAAAGGCG..................CGAGGGATAT 476
      ||||   |||||||||   :::                 ::::::
1460 aAlaThrProLeuProThrSerPheProThrLeuSerPheGlySerLeuL 1477

 477 ATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGA 526
       :::::::::|||      :::||||||      ::::::   :::
1477 euSerSerAlaThrThrProSerLeuProMetSerAlaGlyArgSer... 1492

 527 CCGACAACCGCAGCACCGGACAACGGCTTGCCG................. 559
         |||      |||   ::::::::||||||
1493 ...ThrGluGluAlaThrSerSerAlaLeuProGluLysProGlyAspSe 1508

 560 ....ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGA 605
        :::   ::::::      :::|||     :::::::::   |||
1508 rGluValSerAlaSerAlaAlaSerLeuLeuGluGluGlnGlnSerAlaG 1525

 606 CGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCA 655
              :::|||||||||:::|||:::    |||          :::
1525 lnLeuProGlnAlaProProGlnThrSerAspSerValLysLysGluPro 1541

 656 ATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATC 705
     :::        :::|||:::::::   :::     ::::::   :::|||||
1542 ValLeuAlaGlnProAlaValSer...AsnSerGlyThrAlaAlaSerSe 1557

 706 GGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAG 755
      |:::                |||||||||:::|||   |||
1557 rThrSerLeuValAlaLeuSerAlaGluAlaThrPro............. 1569

 756 CGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCG 805
            |||   |||   :::       :::          :::   ||||||
1570 .....AlaThrThrGlyValProAspAlaArgThrGluAlaValProPro 1584

 806 AAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAA 855
                                               :::|||
1585 .......................................AlaSerSerPh 1588

 856 GACTATGCCGCAGCAG.................................. 871
      ::::::|||    |||
1588 eSerValProGlyGlnThrAlaValThrAlaAlaAlaIleSerSerAlaG 1605

 872 ..CCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAG 919
        |||   |||:::        |||:::|||||||:::
1605 lyProValAlaValGluThrSerSerThrProIle............... 1616

 920 AAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGA 969
               |||:::   |||:::       |||:::||||||
1617 AlaSerSerThrThrSerIleValAlaProGlyProSerAlaGluAlaAl 1633

 970 GCTGTCCGGGGAAAATACGGCTTGGGCG.........GCATCACGGCACA 1010
               |||      |||:::   :::       |||:::
1633 aAlaPheGlyThrValThrSerGlySerSerValPheAlaGlnProProA 1650

1011 TCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAAT 1060
            ||||||:::        :::   :::          :::   ::::::
1650 laAlaSerSerSerSerAlaPheAsnGlnLeuThrAsnAsnThrAlaThr 1666

1061 CCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCC 1110
          |||||||||||   :::      :::  :::    :::|||    ||
1667 AlaProSerAlaThrProValPheGlyGlnValAlaAlaSerThrAlaPr 1683

1111 C.........................................CTTACCA 1118
     |                                             :::
1683 oSerLeuPheGlyGlnGlnThrGlySerThrAlaSerThrAlaAlaAlaT 1700

1119 TTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACA 1168
        |||:::::|||    :::     :::||||||    :::          |||
1700 hrProGlnValSerSerSerGlyPheSerSerProAlaPheGlyThrThr 1716

1169 TCACCT 1174
     :::|||
1717 AlaPro 1718
```

```
seq_name: SwissProt_40:ALP_CEPAC

seq_documentation_block:
ID   ALP_CEPAC      STANDARD;      PRT;   402 AA.
AC   P29118;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
GN   ALP.
OS   Cephalosporium acremonium (Acremonium chrysogenum).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acremonium.
OX   NCBI_TaxID=5044;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91299283; PubMed=1368696;
RA   Isogai T., Fukagawa M., Kojo H., Kohsaka M., Aoki H., Imanaka H.;
RT   "Cloning and nucleotide sequences of the complementary and genomic
RT   DNAs for the alkaline protease from Acremonium chrysogenum.";
RL   Agric. Biol. Chem. 55:471-477(1991).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D00923; BAA00765.1; -.
DR   PIR; JU0332; JU0332.
DR   HSSP; P06873; 2PRK.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21    120       POTENTIAL.
FT   CHAIN       121    402       ALKALINE PROTEINASE.
FT   ACT_SITE    160    160       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    191    191       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    347    347       CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ   SEQUENCE   402 AA;  42099 MW;  8D030CCD42D918E1 CRC64;

alignment_scores:
            Quality:  120.00           Length:    324
              Ratio:  0.745              Gaps:     15
 Percent Similarity:  49.691   Percent Identity: 22.222

alignment_block:
 US-09-303-518D-463 x ALP_CEPAC   ..

 Align seg 1/1  to: ALP_CEPAC  from: 1  to: 402

      76 GATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTT 125
          ::::::   |||:::|||   :::   :::|||
      90 GluIleLysAsnAsnProAspValLeuGluVal................. 100

     126 CGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAG....... 168
            |||   |||       :::|||||||||   ::::::   |||
     101 .GluGluAspGlnIleTrpHisLeuPheAspGluGlnAspGluGlyGluP 117

     169 .............CTTGCCNAGCGCAACGGCCATATCGGATTGGGAAAC 204
                      |||:::::::::||||||       |||||||||:::
     117 heSerThrAlaAlaLeuValThrGlnAsnGlyAlaTrpGlyLeuGlyThr 133
```

```
205 ATACAAAGCCATCAGTTGGGC......CACCTGATGATTCAACAGGCGGC 248
    |||:::  :::|||  |||          :::  :::::::::||
134 IleSerHisArgGlnProGlySerThrSerTyrIleTyrAspAspSerAl 150

249 CGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGG...C 295
    |  :::|||:::  :::|||:::|||           |||  |||
150 aGlySerGlyThrTyrAlaTyrValVal.........AspThrGlyIleL 164

296 ACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGAC 345
      :::   |||:::  |||:::::::::|||     :::  :::
164 euGluSerHisAsnGluPheSerGlyArgAlaIleThrGlyTyrAsnAla 180

346 GAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGA 395
      :::||||||  :::|||                        ::
181 ValGlyGlySerAsnAlaAsp.......................ThrAs 189

396 CGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGC. 444
    :|||:::     |||  |||            |||    ||||||
189 nGlyHisGlyThrHisValAla............GlyThrIleGlyGlyA 202

445 ..GGCTATCCCGCTCCCAAAGGCGCGAGG..................... 471
        |||   :::  |||::::::::
202 rgThrTyrGlyValAlaLysAsnThrAsnLeuIleAlaValLysValPhe 218

472 ................................GATATATACAGCTACGA 488
                                    |||   :::::::::
219 ArgGlySerSerSerSerThrSerIleIleLeuAspGlyPheAsnTrpAl 235

489 CATAAAAGGCGTTGCCCAAAATATCCGCCTCAAC...CTGACCGACAACC 535
     ::::::  :::  ::::::  |||  |||       :::   :::
235 aValAsnAspIleIleAsnArgGlyArgGlnAsnLysAlaAlaIleSerM 252

536 GCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCC......... 576
     |||  |||         ::::::  |||:::||||||
252 etSerLeuGlyGlyGlyTyrSerSerAlaPheAsnAsnAlaValAsnThr 268

577 ...........GGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAA 614
              |||:::::   :::   ::::::|||:::       ::
269 AlaTyrSerArgGlyValLeuSerValValAlaAlaGlyAsnAspAsnGl 285

615 ACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCG 664
    ::::|||::::::|||||||||              ::::::||||||:
285 nAsnAlaAlaAsnTyrSerPro...............AlaSerAlaAlaA 297

665 AAGCC............................................. 669
    ::|||
297 snAlaIleThrValGlySerIleAlaSerAsnTrpAlaArgSerSerPhe 313

670 ......TTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGC 713
          ::::::::::::  ||||||               :::
314 SerAsnTyrGlySerValLeuAspIlePheAlaProGlyThrSerIleLe 330

714 AGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCT 763
     :::        :::|||::::::::|||::::::   ||||||   |||:
330 uSerAlaTrpIleGlyGlyAsnSerAlaThrAsnThrIleSer...GlyT 346

764 CAAACATTGCTGTCATGCACGGCTTGGGTCTG.................. 795
    ::::::::|||:::   |||      |||:::
346 hrSerMetAlaThrProHisValThrGlyValValLeuTyrLeuGlnAla 362

796 .........CTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGC 836
             |||:::|||::::::        |||||||:::|||    |||||
363 LeuGluGlyLeuThrThrSerGlyAlaAlaAlaArgLeuAsnAlaLeuAl 379

837 AGATATGGCGCAACTCAAAGAC 858
    |     ::::::::::::::
379 aThrThrGlyArgValSerAsn 386
```

seq_name: SwissProt_40:MUC4_HUMAN

seq_documentation_block:

```
ID   MUC4_HUMAN     STANDARD;      PRT;   610 AA.
AC   Q99102;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 4 (Tracheobronchial mucin) (Fragment).
GN   MUC4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91207329; PubMed=1673336;
RA   Porchet N., van Cong N., Dufosse J., Audie J.P., Guyonnet-Duperat V.,
RA   Gross M.S., Denis C., Degand P., Bernheim A., Aubert J.P.;
RT   "Molecular cloning and chromosomal localization of a novel human
RT   tracheo-bronchial mucin cDNA containing tandemly repeated sequences
RT   of 48 base pairs.";
RL   Biochem. Biophys. Res. Commun. 175:414-422(1991).
CC   -!- TISSUE SPECIFICITY: TRACHEO-BRONCHIAL AND COLONIC MUCOSAE.
CC   -!- PTM: HIGHLY O-GLYCOSYLATED.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64594; AAA63230.1; -.
DR   PIR; PN0012; PN0012.
DR   MIM; 158372; -.
KW   Repeat; Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN        1    608       38 X 16 AA APPROXIMATE TANDEM REPEATS.
FT   NON_TER     610    610
SQ   SEQUENCE   610 AA;  58670 MW;  52EC24347D31BCEE CRC64;
```

alignment_scores:

```
         Quality:  120.00         Length:    416
           Ratio:  0.632            Gaps:     21
 Percent Similarity:  45.673  Percent Identity:  22.837
```

alignment_block:
US-09-303-518D-463 x MUC4_HUMAN ..

Align seg 1/1 to: MUC4_HUMAN from: 1 to: 610

```
113 ACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGG 162
    |||   :::   |||        ||||||:::||||||
159 ThrAspAlaSerSerAlaSerThrGlyHisThrThr.............. 170

163 GGGGAGCTTGCCN......AGCGCAACGGCCATATCGGATTGGGAAACAT 206
        |||||||||       :::|||:::::::::|||       :::|||
171 ....SerLeuProValThrAspAlaSerSerValSerThrGlyAspThr. 185

207 ACAAAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAG 256

185 .................................................. 185

257 GAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCAT 306
                      :::   :::|||:::|||   |||:::|||
186 .................ThrProLeuProValThrAspThrSerSerAl 196

307 TCGCCCTTCGACAACCATGCCT...CACATT............CCGATTC 341
          :::   |||   |||   |||:::                 :::
196 aSerThrGlyAspThrThrProLeuHisValThrAspAlaSerSerValS 213

342 TGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATT 391
      |||       ::::::||||||     ::::::::::::::
213 erThrGlyHisAlaThrProLeuHisValThrSerLeuSerSerValSer 229
```

```
 392 GGGACGGATACGAACACCATCCCGCCGACG............GCTATGAC 429
         ||||||:::   :::|||   |||           |||   ::
 230 ThrGlyAspThrThrProLeuProValThrSerProSerSerAlaSerSe 246

 430 GGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATA 479
     :||||||   |||::::::::|||:::        |||   ::::::  |
 246 rGlyHis...AlaThrSerLeuProValThrAspAlaSerSerValSerT 262

 480 CA......GCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACC 523
     ||      ||||||:::          :::         |||||||||:::
 262 hrGlyHisAlaThrSerLeuProValThrIleProSerSerAlaSerSer 278

 524 TGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAAT 573
           :::|||:::   |||   ::::::|||   ::::::||||||
 279 GlyAspAlaThrSerLeuProValThrSerLeuSerSerLeuSerThrGl 295

 574 GCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCAC 623
       |||
 295 yHisAla........................................... 297

 624 CCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCA 673
                                         |||   |||
 298 .................................ThrProLeuProVal 302

 674 ACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATT 723
     |||:::|||:::   :::|||
 303 ThrSerLeuSerSerAlaSer............................. 309

 724 GTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGC 773
      :::::::|||   |||      |||   ::::::   |||:::::::::
 310 .ThrGlyHisAlaThrProLeuProVal..ThrAspThrSerSerValSe 325

 774 TGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCG.... 819
     ::::   |||:::   :::||||||   |||:::   :::   :::
 325 rThrGlyHisAlaThrSerLeuLeuValThrAspAlaSerSerValSerT 342

 820 ...................CGCATCAACGATTTGGCAGATATGGCGCAA 849
                        :::::::::|||   :::::::::::: ::
 342 hrGlyHisAlaThrProLeuHisValThrAspAlaSerSerValSer.Th 358

 850 CTCAAAGACTATGCCG...........CAGCAGCCATCCGCGATTGGGC 887
     ::::   |||   |||            :::      ||||||
 358 rGlyAspThrThrProLeuProValThrSerProSerSerAlaSerThrG 375

 888 AGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT 937
       :::   ||||||:::|||          :::        |||::::::|||
 375 lyAspThrThrProLeuPro......ValThrAspThrSerSerValSer 389

 938 TTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATAC 987
           :::   :::|||           :::   |||:::   :::||
 390 ThrGlyAspThrThrProLeuLeuValThrAspThrSerSerValSerTh 406

 988 GGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGC 1037
     |:::   |||:::|||      :::   :::::::::         :::
 406 rSerHisAlaThrSerLeuProValThrAspThrSer...SerValSerT 422

1038 GATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATG 1087
      ||||||                         ||||||   ||||||:::
 422 hrSerHis......................AlaThrSerLeuProVal 430

1088 ...............CGGCATACGCCAAATACCCGTCCCCTTACCATTCC 1122
                     |||   :::|||   ||||||   :::
 431 ThrAspProSerSerAlaSerThrGlyAspThrThrProLeuProValTh 447

1123 CGAAATATCC...................................GTTC 1136
      :::   |||                                   ::::
 447 rAspThrSerSerValSerThrSerHisAlaThrSerLeuProValThrA 464

1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGC 1186
     ::|||   |||:::   |||:::      ||||||   |||
 464 spThrSerSerAlaSerThrSerHisAlaThrSerLeuProVal...... 478

1187 CGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCA 1229
       :::|||::::::   |||  ||||||:::   ||||||
 479 ...ThrAspThrSerSerAlaSer.ThrGlyHisAlaThrPro 491
```

seq_name: SwissProt_40:DAG1_MOUSE

```
seq_documentation_block:
ID   DAG1_MOUSE     STANDARD;      PRT;   893 AA.
AC   Q62165; Q61497; Q61141; Q61094;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dystroglycan precursor (Dystrophin-associated glycoprotein 1)
DE   [Contains: Alpha-dystroglycan (Alpha-DG); Beta-dystroglycan (Beta-
DE   DG)].
GN   DAG1 OR DAG-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   STRAIN=129/SVJ;
RX   MEDLINE=97318791; PubMed=9175728;
RA   Williamson R.A., Henry M.D., Daniels K.J., Hrstka R.F., Lee J.C.,
RA   Sunada Y., Ibraghimov-Beskrovnaya O., Campbell K.P.;
RT   "Dystroglycan is essential for early embryonic development:
RT   disruption of Reichert's membrane in Dag1-null mice.";
RL   Hum. Mol. Genet. 6:831-841(1997).
RN   [2]
RP   SEQUENCE OF 1-650 FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=97210754; PubMed=9057818;
RA   Brancaccio A., Ruegg M.A., Engel J.;
RT   "Cloning and sequencing of mouse skeletal muscle alpha-dystroglycan.";
RL   Matrix Biol. 14:681-685(1995).
RN   [3]
RP   REVISIONS TO 142-143.
RA   Brancaccio A.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 352-650 FROM N.A., AND TISSUE SPECIFICITY.
RC   STRAIN=C57BL/10; TISSUE=Skeletal muscle;
RX   MEDLINE=95135414; PubMed=7833916;
RA   Gorecki D.C., Derry J.M.J., Barnard E.A.;
RT   "Dystroglycan: brain localisation and chromosome mapping in the
RT   mouse.";
RL   Hum. Mol. Genet. 3:1589-1597(1994).
RN   [5]
RP   SEQUENCE OF 620-893 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Decidua;
RX   MEDLINE=97026286; PubMed=8872465;
RA   Yotsumoto S., Fujiwara H., Horton J.H., Mosby T.A., Wang X.,
RA   Cui Y., Ko M.S.H.;
RT   "Cloning and expression analyses of mouse dystroglycan gene: specific
RT   expression in maternal decidua at the peri-implantation stage.";
RL   Hum. Mol. Genet. 5:1259-1267(1996).
RN   [6]
RP   DISULFIDE BOND.
RX   MEDLINE=99116369; PubMed=9917844;
RA   Brancaccio A., Jeno P., Engel J.;
RT   "A single disulfide bridge (Cys182-Cys264) is crucial for alpha-
RT   dystroglycan N-terminal domain stability.";
RL   Ann. N.Y. Acad. Sci. 857:228-231(1998).
CC   -!- FUNCTION: FORMS PART OF THE DYSTROPHIN-ASSOCIATED PROTEIN COMPLEX
CC       (DAPC) WHICH MAY LINK THE CYTOSKELETON TO THE EXTRACELLULAR
CC       MATRIX. ALPHA-DYSTROGLYCAN FUNCTIONS AS A LAMININ RECEPTOR. BINDS
CC       TO SEVERAL TYPES OF ARENAVIRUSES. IS A TARGET FOR THE ENTRY OF
CC       MYCOBACTERIUM LEPRAE INTO PERIPHERAL NERVE SCHWANN CELLS.
CC   -!- SUBCELLULAR LOCATION: ALPHA-DYSTROGLYCAN IS A EXTRACELLULAR
CC       PROTEIN WHILE BETA-DYSTROGLYCAN IS A TYPE-I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES. IN BRAIN,
CC       EXPRESSED IN THE HIPPOCAMPAL FORMATION, THE OLFACTORY BULB, THE
```

```
CC       CEREBELLUM AND THE THALAMUS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U48854; AAA99779.2; -.
DR   EMBL; X86073; CAA60031.1; -.
DR   EMBL; Z34532; CAA84293.1; -.
DR   EMBL; U43512; AAC52853.1; -.
DR   MGD; MGI:101864; Dag1.
KW   Signal; Glycoprotein; Transmembrane; Cytoskeleton.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    651       ALPHA-DYSTROGLYCAN.
FT   CHAIN       652    893       BETA-DYSTROGLYCAN.
FT   DOMAIN      652    751       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    752    772       POTENTIAL.
FT   DOMAIN      773    893       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      316    475       PRO-RICH.
FT   DOMAIN      807    893       PRO-RICH.
FT   DISULFID    180    262
FT   DISULFID    667    711       POTENTIAL.
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    647    647       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    659    659       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    448    450       TKK -> SKE (IN REF. 4).
FT   CONFLICT    599    600       GD -> PH (IN REF. 4).
FT   CONFLICT    643    643       I -> V (IN REF. 2 AND 4).
SQ   SEQUENCE   893 AA;  96904 MW;  59C081EA86AB0AC1 CRC64;

alignment_scores:
          Quality:  119.50            Length:    536
            Ratio:  0.515               Gaps:     28
 Percent Similarity:  43.284   Percent Identity: 23.694

alignment_block:
 US-09-303-518D-463 x DAG1_MOUSE   ..

 Align seg 1/1  to: DAG1_MOUSE  from: 1  to: 893

     30 TCTGTCCATACTGGCAGTGTGCCT......GCCGATGCATGCACACGCCT 73
         |||:::::   :::|||||||||          :::   |||
    263 SerLeuAsnGlnAsnSerValProAspIleArgGlyValGluThrProAl 279

     74 CAGATTTGGCAAACGATCCCTT.........TATCCG............ 101
          |||  |||        |||           ||||||
    279 aArgGluGlyAlaMetSerAlaGlnLeuGlyTyrProValValGlyTrpH 296

    102 .....GCAGGTTCTCGACCGTCA...GCATTTCGAACCCGACGGGAAATA 143
             |||::::::::|||:::          |||   ||||||:::|||
    296 isIleAlaAsnLysLysProThrLeuProLysArgLeuArgArgGlnIle 312

    144 ...........CCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCG.. 179
                      |||:::          |||       :::|||
    313 HisAlaThrProThrProValThrAlaIleGlyProProThrThrAlaIl 329

    180 ....CAACGGCCATATCGGATT...GGGAAACATACAAAGCCATCAGTTG 222
            :::   |||   ||||||          |||:::|||:::::::
    329 eGlnGluProProSerArgIleValProThrProThrSerProAlaIleA 346

    223 GGCCACCTGATGATTCAACAGGCGG.CCGTTGAAGGAA........... 259
         ::||||||   ::::::   :::   |||:::::::::
    346 laProProThrGluThrMetAlaProProValArgAspProValProGly 362

    260 ...............................ATATCGGCTACATTGTC 276
                                       |||      |||    ::
    363 LysProThrValThrIleArgThrArgGlyAlaIleIleGlnThrProTh 379

    277 CGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGC 326
         :        ||||||     |||:::      :::   :::||||||:::|
    379 rLeuGlyProIleGlnProThrArgValSerGluAlaGlyThrThrValP 396

    327 CT....................CACATTCCGATTCTGACGAAGCCGGTA 355
         ||                        ||||||       :::::|||:::
    396 roGlyGlnIleArgProThrLeuThrIleProGlyTyrValGluProThr 412

    356 GTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAA 405
         :::   :::|||
    413 AlaValIleThr..................................... 416

    406 CACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGC 455
                ||||||||:::    |||            ::::::   |||
    417 .......ProProThrThrThrThrLysLysProArgValSerThrProL 431

    456 TCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCC 505
           |||   :::    :::      ::::::||||||       |||    ||
    431 ysProAlaThrProSerThrAspSerSerThrThr.ThrThrArgArgPr 447

    506 AAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTT 555
         |           |||:::        |||   |||        |||:::|||:::
    447 oThrLysLysProArgThrProArg...ProValProArgValThrThrL 463

    556 GCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGA 605
              |||:::                 |||         |||
    463 ysAlaProIle.........ThrArgLeuGluThrAlaSerProProThr 476

    606 CGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCA 655
         ||||||:::|||                     ||||||:::|||:::    :::||
    477 ArgIleArgThrThrThrSerGlyValProArgGlyGlyGluProAsnGl 493

    656 ATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATC 705
         |     |||    :::|||::::::|||      |||     |||            :::
    493 n...ArgProGluLeuLysAsnHis.IleAspArgValAspAlaTrpVal 508

    706 GGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAG 755
         |||:::        |||:::             :::|||:::
    509 GlyThrTyrPheGluValLysIleProSerAspThrPheTyr........ 522

    756 CGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCG 805
                                                          :
    523 .................................................A 523

    806 AAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAA 855
         :::|||:::     :::         :::     |||             :::|||:::
    523 spAsnGluAspThrThrThrAspLysLeuLysLeuThrLeuLysLeuArg 539

    856 GACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGC 905
         :::           ::::::   ::::::|||:::        |||    |||:::
    540 GluGlnGlnLeuValGlyGluLysSerTrpValGlnPheAsnSerAsnSe 556

    906 C......GCACAAGGCATAGAAGCCGTCAGCAATATC............ 936
         :                |||:::           |||::::::
    556 rGlnLeuMetTyrGlyLeuProAspSerSerHisValGlyLysHisGluT 573

    937 ..TTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAA 984
              ||||||   |||               |||::::::||||||
    573 yrPheMetHisAlaThrAspLysGlyGlyLeuSerAlaVal......... 586

    985 TACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGG 1034
                    ::::::         |||        ||||||    |||    ||
    587 .........AspAlaPheGluIleHisValHisLysArgProGln...Gl 599

   1035 CGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCG 1084
         |        |||       :::     |||:::     ::::::     :::    |||
    599 yAspLysAlaProAlaArgPheLysAlaArgLeuAlaGlyAspProAlaP 616

   1085 AT........................GCGGCATACGCCAAATACCCG 1107
                                         |||           |||
    616 roValValAsnAspIleHisLysLysIleAlaLeuValLysLysLeuAla 632
```

```
1108 TCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGG 1157
          :::    :::||||||        |||        |||:::
 633 PheAlaPheGlyAspArgAsnCysSerSerIleThrLeuGlnAsnIleTh 649

1158 CAAAGAAAACATC...........ACCTCCTCAACCGTGCCG....... 1188
       :::    :::|||             |||:::::|||:::|||
 649 rArgGlySerIleValValGluTrpThrAsnAsnThrLeuProLeuGluP 666

1189 .....CCGTCAAAC.........GGCAAAAATGTCAAACTGGCAGACCAA 1224
          |||:::::         |||    :::    ::::::|||||||:::
 666 roCysProLysGluGlnIleIleGlyLeuSerArgArgIleAlaAspGlu 682

1225 CGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTT 1274
                               :::||||||      |||    ||
 683 ..........................AsnGlyLysProArgProAlaPh 690

1275 TGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGG 1324
      |::::::    ::::::    |||    |||              ::::::|
 690 eSerAsnAlaLeuGluProAspPheLysAlaLeuSerIleAlaValThrG 707

1325 GGGGCGGT....................ATACCTAAGGCT......AAG 1347
      ||:::|||                    ||||||    |||        :::
 707 lySerGlySerCysArgHisLeuGlnPheIleProValAlaProProSer 723

1348 CCTGTGTTTGATGCGAAACCGAGATGGGAGGTT...GATAGGAAGCTTAA 1394
      |||        :::|||    |||       ||||||   ||||||       ::
 724 ProGlySerSerAlaAlaProAlaThrGluValProAspArgAspProGl 740

1395 TAAA 1398
      :|||
 740 uLys 741
```

seq_name: SwissProt_40:CA1E_CHICK

```
seq_documentation_block:
ID   CA1E_CHICK     STANDARD;      PRT;  1888 AA.
AC   P32018;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(XIV) chain precursor (Undulin).
GN   COL14A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=93185668; PubMed=8444186;
RA   Waelchli C., Trueb J., Kessler B., Winterhalter K.H., Trueb B.;
RT   "Complete primary structure of chicken collagen XIV.";
RL   Eur. J. Biochem. 212:483-490(1993).
RN   [2]
RP   SEQUENCE OF 402-1549 FROM N.A.
RX   MEDLINE=92339443; PubMed=1339349;
RA   Trueb J., Trueb B.;
RT   "Type XIV collagen is a variant of undulin.";
RL   Eur. J. Biochem. 207:549-557(1992).
RN   [3]
RP   STRUCTURE BY NMR OF 1852-1885.
RX   MEDLINE=99280705; PubMed=10350466;
RA   Montserret R., Aubert-Foucher E., McLeish M.J., Hill J.M., Ficheux D.,
RA   Jaquinod M., van der Rest M., Deleage G., Penin F.;
RT   "Structural analysis of the heparin-binding site of the NC1 domain of
RT   collagen XIV by CD and NMR.";
RL   Biochemistry 38:6479-6488(1999).
CC   -!- FUNCTION: AN ADHESIVE ROLE BY INTEGRATING COLLAGEN BUNDLES. IT IS
CC       PROBABLY ASSOCIATED WITH THE SURFACE OF INTERSTITIAL COLLAGEN
CC       FIBRILS VIA COL1. THE COL2 DOMAIN MAY THEN SERVE AS A RIGID ARM
CC       WHICH STICKS OUT FROM THE FIBRIL AND PROTRUDES THE LARGE N-
CC       TERMINAL GLOBULAR DOMAIN INTO THE EXTRACELLULAR SPACE, WHERE IT
CC       MIGHT INTERACT WITH OTHER MATRIX MOLECULES OR CELL SURFACE
CC       RECEPTORS.
CC   -!- SUBUNIT: HOMOTRIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC   -!- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION; HIGH PRESENCE IN
CC       DENSE CONNECTIVE TISSUE IN SKELETAL MUSCLE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
CC   -!- PTM: MAY CONTAIN CYSTEINE RESIDUES INVOLVED IN INTER- AND
CC       INTRAMOLECULAR DISULFIDE BONDING.
CC   -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC       INTERRUPTED HELICES (FACIT) FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X70793; CAA50064.1; -.
DR   EMBL; X70792; CAA50063.1; -.
DR   EMBL; X66138; CAA46928.1; -.
DR   PIR; S22916; S22916.
DR   PDB; 1B9P; 25-FEB-99.
DR   PDB; 1B9Q; 25-FEB-99.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF01391; Collagen; 4.
DR   Pfam; PF00041; fn3; 8.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00092; vwa; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   SMART; SM00060; FN3; 7.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00327; VWA; 2.
DR   PROSITE; PS50234; VWFA; 2.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cell adhesion; Collagen; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   1888       COLLAGEN ALPHA 1(XIV) CHAIN.
FT   DOMAIN      158    330       VWFA 1.
FT   DOMAIN      441    529       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      531    622       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      625    728       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      739    821       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      829    922       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      952   1010       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1042   1215       VWFA 2.
FT   DOMAIN     1227   1468       NONHELICAL REGION (NC4).
FT   DOMAIN     1469   1620       TRIPLE-HELICAL REGION 1.
FT   DOMAIN     1664   1786       TRIPLE-HELICAL REGION 2.
FT   DOMAIN      721    733       POLY-THR.
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1398   1398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE       1489   1491       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1617   1619       CELL ATTACHMENT SITE (POTENTIAL).
SQ   SEQUENCE   1888 AA;  202666 MW;  39915BB9F46DD873 CRC64;

alignment_scores:
          Quality:  119.50          Length:    708
            Ratio:  0.373             Gaps:     42
```

```
Percent Similarity:  45.198   Percent Identity: 21.610

alignment_block:
 US-09-303-518D-463 x CA1E_CHICK   ..

 Align seg 1/1  to: CA1E_CHICK  from: 1  to: 1888

  83 CAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 132
      :::   :::   |||||||||     ::::::     |||:::::::
 667 GluGluValValLeuSerGlyAspAlaAspSerTyrValIleGluGlyLe 683

 133 GACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAA 182
             ||||||   |||::::::                 |||
 683 uLeuProAsnThrGluTyrGluVal.................SerLeuL 694

 183 CGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGA 232
       |||:::   |||   ||||||   :::    :::
 694 euAlaValPheAspAspGluThrGluSerGluVal.............. 705

 233 TGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTG........ 274
                        :::      ::::::||||||:::
 706 .................ValAlaValLeuGlyAlaThrIleValGlyTh 716

 275 .TCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCA 323
      :::|||:::|||   |||   |||:::::::         :::::::|||
 716 rThrAlaIleProThrThrValThrThrThrThrThrThrThrAlaThrT 733

 324 TGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCG........... 361
        |||                     ||||||::::::
 733 hrPro.................LysProThrIleAlaValPheArgThr 743

 362 ...............TTGACGGATTCAGCCTTTACCGCATCCATTGGG.. 394
                     :::    |||:::      |||:::|||:::
 744 GlyValArgAsnLeuValIleAspAspGluThrThrSerSerLeuArgVa 760

 395 .............ACGGATACGAACACC...........ATCCCGCCGA 419
                   :::|||    |||:::           :::
 760 lValTrpAspIleSerAspHisAsnAlaGlnGlnPheArgValThrTyrL 777

 420 CGGCTATGACGGGCCACAGGGCG...GCGGCTATCCCGCTCCCAAAGGCG 466
        :::      |||   ||||||      ||||||    :::|||    ||
 777 euThrAlaLysGlyAspArgAlaGluGluAlaIleMetValPro.GlyAr 793

 467 CGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGC 516
     |::::::   :::   |||:::                 :::|||  :::
 793 gGlnAsnThrLeuLeuLeuGlnProLeuLeuProAspThrGluTyr.Lys 809

 517 CTCAACCTGACCGACAACCGCAGCACCGGACAA................ 549
     :::::::::|||         :::    |||:::
 810 ValThrIleThrProIleTyrAlaAspGlyGluGlyValSerValSerAl 826

 550 ........................................CGGCTTGCCG 559
                                            |||::::::|
 826 aProGlyLysThrLeuProLeuSerAlaProArgAsnLeuArgValSerA 843

 560 ACCGTTTCCACAAT....................GCCGGCGCTATGCTG 588
     ||:::::::::|||                        |||
 843 spGluTrpTyrAsnArgLeuArgIleSerTrpAspAlaProProSerPro 859

 589 ACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGA 638
     |||:::|||                :::|||                ::
 860 ThrMetGlyTyrArgIleValTyrLys.....................Se 869

 639 GCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGC.........A 679
     :::::::       |||   |||   |||:::|||   |||         :
 869 rIleAsnValProGlyProAlaLeuGluThrPheValGlyAspAspIleA 886

 680 CTGCAGATATCGTCAAAAACATCATCGGCGCG.................. 711
     :::::    ::::::     |||::::::::::::
 886 snThrIleLeuIleLeuAsnLeuPheSerGlyThrGluTyrSerValLys 902

 712 ......GCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAG 755
           |||:::      :::|||    :::||||||:::    |||:::::

 903 ValPheAlaSerTyrSerThrGlyPheSerAspAlaLeuThrGlyValAl 919

 756 CGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCG 805
     ::::               :::::::::    |||||||:::   :::   :
 919 aLys..............ThrLeuTyr...LeuGlyValThrAsnLeuA 930

 806 AAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAA 855
     :::::    ::::::|||::::::::::|||:::     :::    ||||||
 930 spThrTyrGlnValArgMetThrSerLeuCysAlaGlnTrpGlnLeuHis 946

 856 GACTATGCCGCAGCA..................GCCATCCGCGATTGGGC 887
        :::|||:::|||                   ::::::    |||
 947 ArgHisAlaThrAlaTyrArgValValIleGluSerLeuValAspGlyLy 963

 888 AGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT 937
           |||:::   |||   :::   |||:::                :::    |
 963 sLysGlnGluValAsnLeuGlyGlyGlyValPro......ArgHisCysP 978

 938 TTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATAC 987
     ||:::      :::|||                              :::|||
 978 hePheGluLeuMetPro...................GlyThrGluTyr 987

 988 GGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGC 1037
                  |||::::::|||              :::|||:::
 988 .........LysIleSerValHis...........AlaGlnLeuGlnGl 997

1038 GATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATG 1087
      |||        |||    :::    |||     ::::::  |||    |||||||:::
 997 uIleGluGlyPro...AlaValSerIleMetGlu.ThrThrLeuProPhe 1012

1088 CGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCA 1137
              |||     |||     |||    |||
1013 ProThrGlnProProThrSerProSerThr.................... 1022

1138 AACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCA.ACCGTGC 1186
                                     |||    ||||||    |||:::|
1023 ..........................ThrLeuProProProThrIleP 1030

1187 CGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAG 1236
     ||||||:::          |||:::|||         |||               |||
1030 roProAla......LysGluValCysLysAla...........AlaLys 1040

1237 ACAGGCGTACCGTTT......GACGGTAAA.........GGGTTTCCGAA 1271
     :::    :::    |||         |||||||:::          |||      ||
1041 AlaAspLeuValPheLeuValAspGlySerTrpSerIleGlyAspAspAs 1057

1272 TTTTGAGAAGCACGTGAAA......TATGATACGAAGCTCGATATTCAAG 1315
     ||||:::|||     ::::::       |||:::|||          :::::::
1057 nPheAsnLysIleIleSerPheLeuTyrSerThrValGlyAlaLeuAspL 1074

1316 AATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAA 1365
     :::::::        |||          |||         |||:::
1074 ysIleGlyProAspGlyThrGlnValAlaIleIleGlnPheSerAspAsp 1090

1366 CCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGCA 1415
     ||||||    |||            |||||||||          |||:::|||
1091 ProArgThrGluPhe......LysLeuAsnAlaTyrLysThrLysGluTh 1105

1416 GGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGT 1465
      :::    :::    :::|||:::
1105 rLeuLeuGluAlaIleGlnGlnIle........................ 1113

1466 TTAAAGCCCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATTTT 1515
                  |||    :::      ::::::|||||||||         :::
1114 ...........AlaTyrLysGlyGlyAsnThrLysThrGlyLysAlaIle 1126

1516 AATCAT...............TTTATAGGTGGTGATATCAATAAGAAAGG 1550
     :::|||                     |||    |||              :::||||||
1127 LysHisAlaArgGluValLeuPheThrGlyGluAlaGlyMetArgLysGl 1143

1551 C.................ACAGTAACAGGAGGGCATAGTCTAACCCGTG 1582
     |                    ::::::|||    |||:::|||          :::
1143 yIleProLysValLeuValValIleThrAspGlyArgSer......GlnA 1158
```

```
1583 GTGATGTACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTT 1632
         ||||||:::   ::::::::          |||     :::
1158 spAspValAsnLysValSerArgGluMetGlnLeuAspGlyPheSerPhe 1174

1633 TATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTGAA 1682
      :::          |||            |||:::::::::::
1175 PheAlaIleGlyVal...........AlaAspAlaAspTyrSerGluLe 1187

1683 AACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAG 1732
       ::::::   |||:::|||          :::|||:::::|||      |
1187 uValAsnIleGlySerLysProSerGluArgHisValPhePheValAspA 1204

1733 ATTGGGATGAG...GCTAGAATTAGGGCTGAAGTT.............. 1764
      ||:::|||       ::::::|||:::   |||:::
1204 spPheAspAlaPheThrLysIleGluAspGluLeuIleThrPheValCys 1220

1765 ...ACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGATAATAAATGGCA 1811
         |||::::::          :::::::::||||||
1221 GluThrAlaSerAlaThrCysProLeuValPheLysAsp........... 1233

1812 GGGTACAAGTAAATCGGGTATTAAA........................ 1836
      |||    :::    :::|||:::|||
1234 .GlyAspLysLeuAlaGlyPheLysMetMetGluMetPheGlyLeuValG 1250

1837 ................ATAGAAGGATTTACC...GAACCTAATAGA... 1863
                     |||:::|||   :::   |||||||:::
1250 luLysGluPheSerAlaIleAspGlyValSerMetGluProGlyThrPhe 1266

1864 ACAGCATATCCCATTTAT 1881
      ::::::||||||   |||
1267 AsnValTyrProCysTyr 1272
```

seq_name: SwissProt_40:P1P_LACLC

```
seq_documentation_block:
ID   P1P_LACLC      STANDARD;      PRT;  1902 AA.
AC   P16271;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
DE   proteinase).
GN   PRTP.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG   Plasmid pWV05.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WG2;
RX   MEDLINE=88149035; PubMed=3278687;
RA   Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;
RT   "Nucleotide sequence of the cell wall proteinase gene of
RT   Streptococcus cremoris Wg2.";
RL   Appl. Environ. Microbiol. 54:231-238(1988).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC       SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC       E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC       PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC       ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC       INSULIN B-CHAIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M24767; AAA17677.1; -.
DR   HSSP; P00782; 1S01.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW   Transmembrane.
FT   SIGNAL        1     33
FT   PROPEP       34    187
FT   CHAIN       188   1902       PI-TYPE PROTEINASE.
FT   DOMAIN      188   1876       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1877   1895       MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN     1896   1902       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE    217    217       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    281    281       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    620    620       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DOMAIN     1867   1872       CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                                PROTEINS.
SQ   SEQUENCE   1902 AA;  199910 MW;  2901C7F19B2E5D0B CRC64;

alignment_scores:
          Quality:  119.50             Length:    638
            Ratio:  0.418                Gaps:     35
 Percent Similarity:  44.828    Percent Identity: 19.749

alignment_block:
 US-09-303-518D-463 x P1P_LACLC   ..

 Align seg 1/1  to: P1P_LACLC  from: 1  to: 1902

     271 ATTGTCCGCTTTTCCGAT.........CACGGGCACAAATTCCATTCG.. 309
          :::   :::|||:::|||          ||||||:::   |||:::|||
     242 ValGluLysPheThrAspThrAlaLysHisGlyArgTyrPheAsnSerLy 258

     310 ....CCCTTCGAC...AACCATGCCTCACATTCCGAT......TCTGACG 346
            |||:::       |||:::|||:::::::::|||       :::|||:
     258 sValProTyrGlyPheAsnTyrAlaAspAsnAsnAspThrIleThrAspA 275

     347 AAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGAC 396
          :::::       ||||||
     275 spThr.........ValAsp............................. 278

     397 GGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGG 446
               |||:::|||                  |||   :::   :::||
     279 ......GluGlnHis.................GlyMetHisValAlaGl 287

     447 CTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAG 496
          |     :::    :::|||:::   |||       :::    ::::::    |
     287 yIleIleGlyAlaAsnGlyThrGlyAspAspProAlaLysSerValValG 304

     497 GCGTTGCCCAAAATATCCGCCTC.....................AACCTG 525
          ||||||||   :::    :::|||                     |||
     304 lyValAlaProGluAlaGlnLeuLeuAlaMetLysValPheThrAsnSer 320

     526 ACCGACAACCGCAGCACCGGA..............CAACGGCTTGCCGA 560
             :::    :::||||||                :::   :::   ||
     321 AspThrSerAlaThrThrGlySerSerThrLeuValSerAlaIleGluAs 337

     561 CCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGAT 610
```

```
          |                :::::   :::|||:::::::::::|||:::
 337 pSerAlaLysIleGlyAlaAspValLeuAsnMetSerLeuGlySerAspS 354

 611 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660
           :::   |||      :::|||||||||          ||||||
 354 erGlyAsnGlnThrLeuGluAspProGluLeuAlaAlaValGlnAsnAla 370

 661 GCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGC 710
        |||      :::|||||||||  ::::::                 ::
 371 AsnGlu......SerGlyThrAlaAlaValIle...............Se 380

 711 GGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATA....... 753
     :||||||:::             :::|||:::|||:::::|||:::
 380 rAlaGlyAsnSerGlyThrSerGlySerAlaThrGluGlyValAsnLysA 397

 754 .............................................AGC 756
                                                  |||
 397 spTyrTyrGlyLeuGlnAspAsnGluMetValGlyThrProGlyThrSer 413

 757 GAAGGCTCAAAC...................................... 768
     :::|||:::::
 414 ArgGlyAlaThrThrValAlaSerAlaGluAsnThrAspValIleThrGl 430

 769 ....ATTGCTGTCATGCACGGCTTGGGTCTG.................. 795
        :::::::::      |||    ||||||
 430 nAlaValThrIleThrAspGlyThrGlyLeuGlnLeuGlyProGlyThrI 447

 796 .....CTTTCCACC.........................GAAAACAAGATG 816
          |||||:::                          ::::::|||:::
 447 leGlnLeuSerSerAsnAspPheThrGlySerPheAspGlnLysLysPhe 463

 817 GCGCGCATCAACGATTTGGCA.........GATATGGCGCAACTCAAAGA 857
         ::::::|||   :::        :::   :::   |||   ||
 464 TyrValValLysAspAlaSerGlyAsnLeuSerLysGlyAlaLeuAlaAs 480

 858 CTATGCCGCAGCAGCCATCCGCGATTGGGCAGTC................ 891
      ||||:::|||    |||         |||:::
 480 pTyrThrAlaAspAlaLysGlyLysIleAlaIleValLysArgGlyGluL 497

 892 .........................CAAAACCCCAATGCCGCACAAGGC 915
                              :::     :::|||    |||
 497 euSerPheAspAspLysGlnLysTyrAlaGlnAlaAlaGlyAlaAlaGly 513

 916 ATAGAAGCCGTCAGCAAT.............................AT 935
     :::        |||:::|||
 514 LeuIleIleValAsnAsnAspGlyThrAlaThrProValThrSerMetAl 530

 936 CTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAAT 985
       :::    :::::::::|||      |||:::::::::|||   |||:::
 530 aLeuThrThrThrPheProThrPheGlyLeuSerSerValThrGlyGlnL 547

 986 ACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGC 1035
           :::      :::||||||||||||          :::::::::|||
 547 ysLeuValAspTrpValThrAlaHisPro......AspAspSerLeuGly 561

1036 .........GCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAA 1076
               |||:::::::|||      :::::::::   ::::::|||:::
 562 ValLysIleAlaLeuThrLeuValProAsnGlnLysTyrThrGluAspLy 578

1077 TTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAA 1126
      :::::::|||       :::::::::|||    :::|||    :::   :
 578 sMetSerAsp......PheThrSerTyr...GlyProValSerAsnLeuS 592

1127 ATATCCGTTCAAACTTG.............................GAG 1146
      :::::::   ::::::                              :::
 592 erPheLysProAspIleThrAlaProGlyGlyAsnIleTrpSerThrGln 608

1147 CAGCGTTACGGCAAAGAAAACATC......ACCTCCTCAACCGTGCCGCC 1190
      ::::::   |||      |||:::      ||||||   :::   |||
 609 AsnAsnAsnGlyTyrThrAsnMetSerGlyThrSerMetAlaSerProPh 625

1191 GTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAG 1240
           |||:::::::::   |||   :::      :::   |||:::
 625 eIleAlaGlySerGlnAlaLeuLeuLysGlnAlaLeuAsnAsnLysAsnA 642

1241 GCGTACCGTTT.....................GACGGTAAAGGGTTTCCG 1269
     ::   ||||||                       |||    ::::::
 642 sn...ProPheTyrAlaTyrTyrLysGlnLeuLysGlyThrAlaLeuThr 657

1270 AATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATT 1319
     :::|||   |||   |||:::   :::|||           |||:::::::::
 658 AspPheLeuLysThrValGluMetAsnThrAlaGlnProIleAsnAspIl 674

1320 ATCG..............................GGGGGCGGTATAC 1336
     ::::                                |||:::|||:::
 674 eAsnTyrAsnAsnValIleValSerProArgArgGlnGlyAlaGlyLeuV 691

1337 CTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTGATAGG 1386
           :::|||   ::::::||||||
 691 alAspValLysAlaAlaIleAspAla........................ 699

1387 AAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAAT........ 1428
                                 :::|||||||||
 700 ............................LeuGluLysAsnProSerTh 706

1429 ....GTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCC 1474
         |||   |||
 706 rValValAlaGlu..................................... 710

1475 ATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCATTTT 1524
                       ::::::      ::::::::::::::   |||
 711 .................AsnGlyTyrProAlaValGluLeuLysAspPhe 721

1525 ATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGAGGG......CA 1568
       :::   |||   :::   |||   |||   |||::::::        ||
 722 ThrSerThrAspLysThrPheLysLeuThrPheThrAsnSerThrThrHi 738

1569 TAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTG 1618
      |:::||||||                    |||:::   |||       |
 738 sGluLeuThr..................TyrGlnMetAspSerAsnThrA 749

1619 ATAAACATGGGGTTTATCAAGCGACAGTGGAA.................. 1650
      ||    ::::::||||||   ::::::::::::
 749 spThrAsnAlaValTyrThrSerAlaThrAspProAsnSerGlyValLeu 765

1651 ...ATTAAAAAGCCTGATGGAAGT..............TGGGAGGTGAA 1682
          ||||||   ||||||:::                  ::::::
 766 TyrAspLysLysIleAspGlyAlaAlaIleLysAlaGlySerAsnIleTh 782

1683 AACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATG...TTCCCAA 1729
       :::       |||      ::::::   :::   |||:::   :::||||
 782 rValProAlaGlyLysThrAlaGlnIleGluPheThrLeuSerLeuProL 799

1730 AAGATTGGGATGAG 1743
      ||::::::|||:::
 799 ysSerPheAspGln 803
```

seq_name: SwissProt_40:CCAA_MOUSE

```
seq_documentation_block:
ID   CCAA_MOUSE     STANDARD;      PRT;  2164 AA.
AC   P97445;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
DE   channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium
DE   channel I) (BI).
GN   CACNA1A OR CACNL1A4 OR CCHA1A OR CACH4 OR CACN3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT TG LEU-601.
RC   STRAIN=DBA/2J;
```

```
RX   MEDLINE=97083572; PubMed=8929530;
RA   Fletcher C.F., Lutz C.M., O'Sullivan T.N., Shaughnessy J.D. Jr.,
RA   Hawkes R., Frankel W.N., Copeland N.G., Jenkins N.A.;
RT   "Absence epilepsy in tottering mutant mice is associated with calcium
RT   channel defects.";
RL   Cell 87:607-617(1996).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC       GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
CC       CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC       ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
CC       IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
CC       DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
CC       GVIA).
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC; MAINLY FOUND IN THE
CC       CEREBELLUM, OLFACTORY BULB, CEREBRAL CORTEX, HIPPOCAMPUS, AND
CC       INFERIOR COLLICULUS. IN THE HIPPOCAMPUS, EXPRESSION OCCURS IN
CC       PYRAMIDAL AND GRANULE NEURONS, AS WELL AS IN INTERNEURONS.
CC       PURKINJE CELLS CONTAIN PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING
CC       A PROMINENT CALCIUM CURRENT IN CEREBELLAR GRANULE CELLS.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- DISEASE: DEFECTS IN CACNA1A ARE THE CAUSE OF A DELAYED-ONSET,
CC       RECESSIVE MEUROLOGICAL DISORDER SEEN IN TOTTERING (TG) MUTANTS,
CC       RESULTING IN ATAXIA, MOTOR SEIZURES AND BEHAVIORAL ABSENCE
CC       SEIZURES RESEMBLING PETIT MAL EPILEPSY (OR ABSENCE EPILEPSY) IN
CC       HUMANS. THERE ARE TWO MORE ALLELES: LEANER (TG(LA)), THAT IS
CC       CHARACTERIZED BY SEVERE ATAXIA AND FREQUENT DEATH PAST WEANING,
CC       BUT NO MOTOR SEIZURES; AND ROLLING NAGOYA (TG(ROL)), THAT PRESENTS
CC       AN INTERMEDIARY PHENOTYPE, THE ATAXIA BEING SOMEWHAT MORE SEVERE
CC       THAT WITH TG, BUT WITHOUT MOTORS SEIZURES. SELECTIVE DEGENERATION
CC       OF CEREBELLAR PURKINJE CELLS HAS BEEN SHOWN FOR ALL THESE TYPES OF
CC       MUTANTS.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U76716; AAC52940.1; -.
DR   MGD; MGI:109482; Cacna1a.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
DR   SMART; SM00384; AT_hook; 1.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation; Disease mutation.
FT   REPEAT       17    317       I.
FT   REPEAT      427    671       II.
FT   REPEAT     1134   1417       III.
FT   REPEAT     1454   1717       IV.
FT   DOMAIN        1     52       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     53     71       S1 OF REPEAT I (POTENTIAL).
FT   DOMAIN       72     90       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     91    108       S2 OF REPEAT I (POTENTIAL).
FT   DOMAIN      109    120       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    121    136       S3 OF REPEAT I (POTENTIAL).
FT   DOMAIN      137    144       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    145    163       S4 OF REPEAT I (POTENTIAL).
FT   DOMAIN      164    182       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    183    202       S5 OF REPEAT I (POTENTIAL).
FT   DOMAIN      203    289       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    290    314       S6 OF REPEAT I (POTENTIAL).
FT   DOMAIN      315    441       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    442    461       S1 OF REPEAT II (POTENTIAL).
FT   DOMAIN      462    475       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    476    495       S2 OF REPEAT II (POTENTIAL).
FT   DOMAIN      496    503       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    504    522       S3 OF REPEAT II (POTENTIAL).
FT   DOMAIN      523    532       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    533    551       S4 OF REPEAT II (POTENTIAL).
FT   DOMAIN      552    570       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    571    590       S5 OF REPEAT II (POTENTIAL).
FT   DOMAIN      591    643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    644    668       S6 OF REPEAT II (POTENTIAL).
FT   DOMAIN      669   1142       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1143   1166       S1 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1167   1183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1184   1203       S2 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1204   1210       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1211   1234       S3 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1235   1245       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1246   1263       S4 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1264   1282       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1283   1302       S5 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1303   1389       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1390   1414       S6 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1415   1470       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1471   1489       S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1490   1503       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1504   1525       S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1526   1532       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1533   1552       S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1553   1559       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1560   1578       S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1579   1597       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1598   1617       S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1618   1689       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1690   1715       S6 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1716   2164       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      681    686       POLY-GLU.
FT   DOMAIN     1107   1110       POLY-GLU.
FT   DOMAIN     2114   2123       POLY-HIS.
FT   DOMAIN      337    354       BINDING TO THE BETA SUBUNIT (BY
FT                                SIMILARITY).
FT   SITE        272    272       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE        622    622       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1363   1363       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1659   1659       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   MOD_RES    1725   1725       PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND    1743   1754       BY SIMILARITY.
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1559   1559       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     601    601       P -> L (IN TG).
SQ   SEQUENCE   2164 AA;  246978 MW;  E9A80AC5B4758BBC CRC64;

alignment_scores:
```

```
           Quality:  119.50          Length:    461
             Ratio:  0.626             Gaps:     27
Percent Similarity:  41.432   Percent Identity: 23.861

alignment_block:
 US-09-303-518D-463 x CCAA_MOUSE   ..

 Align seg 1/1  to: CCAA_MOUSE  from: 1  to: 2164

    183 CGGCCATATCGGATTGGGAAACATACA.........AAGCCATCAGTTGG 223
        ||||||    |||              |||       :::|||   |||
    762 ArgProLeuArgProAspValLysThrHisLeuAspArgProLeuValVa 778

    224 GCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATT 273
            |||::::::::   :::::::::   :::|||
    778 lAspProGlnGluAsnArgAsnAsnAsnThrAsnLys............ 790

    274 GTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCA 323
                     |||||||||      |||    |||||||||
    791 .............SerArgAla......ProGluAlaLeuArg...... 798

    324 TGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCA 373

    798 ................................................. 798

    374 GCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGC 423
          |||      :::|||          |||    |||:::   ||||||
    799 .ProThrAlaArgProArgGluSerAlaArgAspProAspAlaArgArgA 815

    424 TA..............TGACGGGCCACAGGGCGG.............. 443
                        ||||||    ||||||
    815 laTrpProGlySerProGluArgAlaProGlyArgGluGlyProTyrGly 831

    444 CGGCTATCCCGCTCCCAAAGGCGCGAG...................... 470
        |||    |||:::    |||::::||||||
    832 ArgGluSerGluLeuGlnGlnArgGluHisAlaProProArgGluHisAl 848

    471 ................................GGATATATACAGCTAC 486
                                         |||           |
    848 aProTrpAspAlaAspThrGluArgAlaLysAlaGlyAspAlaProArgA 865

    487 GACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCG 536
        |||||    :::|||              |||       :::   :::|||
    865 rgHisThrHisArg...............ProValAlaGluGlyGluPro 876

    537 CAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGC 586
        :::::::::   :::       ||||||        :::
    877 ArgArgHisArgAlaArgArgArgProGlyAspGlu...........Pr 889

    587 TGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCC. 635
          |||    |||    :::||||||    :::     |||         :::|||
    889 oAspAspArgProGluArgArgProArgProArgAspAlaThrArgProA 906

    636 ..CGAGCT...........GGACAGATCGGGCAATGCCGCCGAAGCCTT 671
          ||||||              |||:::   |||:::   :::|||
    906 laArgAlaAlaAspGlyGluGlyAspAspGlyGluArgLysArg...... 920

    672 CAACGGCACTGCAGATAT.....................CGTCAAAAACA 700
           ||||||   |||:::                      |||:::::::::
    921 ...ArgHis...ArgHisGlyProProAlaHisAspAspArgGluArgAr 935

    701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
        :|||||||||    :::::::|||    |||:::|||        |||:::|||
    935 gHisArgArgArgLysGluAsn.GlnGlySerGlyValProValSerGly 951

    751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTC 800
          :::    :::::::::         :::::::::    ||||||
    952 ProAsnLeuSerThrThrArgProIleGlnGlnAspLeuGly........ 965

    801 CACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 850
            ::::::   :::             :::||||||    :::|||  :::|||
    966 ....ArgGlnAspLeuProLeuAlaGluAspLeuAspAsnMet.LysAsn 980

    851 TCAAAGACT......................................... 859
          :::|||
    981 AsnLysLeuAlaThrGlyGluProAlaSerProHisAspSerLeuGlyHi 997

    860 .......ATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAA 902
               :::|||    :::|||:::    ||||||:::|||    :::|||
    997 sSerGlyLeuProProSerProAlaLysIleGlyAsnSerThrAsnProG 1014

    903 TGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCC 952
          |||        |||    :::|||
   1014 lyProAlaLeuAlaThrAsnPro.......................... 1021

    953 CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATC 1002
                                  :::|||         ||||||||
   1022 .........................GlnAsn.........AlaAlaSe 1026

   1003 ACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAA 1052
        ||||                                    |||:::       :::
   1026 rArg.............................ArgThrProAsnAsnP 1033

   1053 AGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAAT 1102
          |||:::|||     :::                |||   :::        |||
   1033 roGlyAspProSerAsnProGly......ProProLysThrProGluAsn 1047

   1103 ACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1152
        :::       :::|||   |||:::    |||:::|||      |||:::
   1048 SerLeuIleValThrAsnProSer...SerThrGlnProAsnSerAlaLy 1063

   1153 TACGGCAAAGAAAACATCA.................CCTCCTCAACCGT 1184
          |||||||:::|||   :::                  ||||||   :::
   1063 sThrAlaArgLysProGluHisMetAlaValGluIleProPro..AlaCy 1079

   1185 GCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGA 1234
          ||||||    |||            |||::::::      :::    :::|||
   1079 sProProLeuAsnHisThrValValGlnValAsnLysAsnAlaAsnProA 1096

   1235 AGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCAC 1284
                                    :::|||:::    |||::::::
   1096 spPro.......................LeuProLysLysGluGluGlu 1104

   1285 GTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGGTAT 1334
           |||    :::    :::    |||                |||
   1105 LysLysGluGluGluGluAlaAsp............ProGlyGluAspGl 1117

   1335 ACCTAAGGCTAAGCCTGTGTTTGATGCG 1362
          ||||||       |||    :::::::::
   1117 yProLysProMetProProTyrSerSer 1126

seq_name: SwissProt_40:ITA4_HUMAN

seq_documentation_block:
ID   ITA4_HUMAN     STANDARD;      PRT;  1038 AA.
AC   P13612;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Integrin alpha-4 precursor (Integrin alpha-IV) (VLA-4) (CD49d).
GN   ITGA4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89356603; PubMed=2788572;
RA   Takada Y., Elices M.J., Crouse C., Hemler M.E.;
RT   "The primary structure of the alpha 4 subunit of VLA-4: homology to
RT   other integrins and a possible cell-cell adhesion function.";
RL   EMBO J. 8:1361-1368(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92201303; PubMed=1551405;
RA   Rubio M., Nueda A., Vara A., Corbi Lopez A.L.;
```

```
RT   "A single mRNA encodes the alpha 150 and alpha 80/70 forms of the
RT   alpha subunit of VLA4.";
RL   Eur. J. Immunol. 22:1099-1102(1992).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANT GLN-884.
RX   MEDLINE=96256895; PubMed=8643114;
RA   Szabo M.C., McIntyre B.W.;
RT   "Identification of two variants of the human integrin alpha 4
RT   subunit.";
RL   Mol. Immunol. 32:1453-1454(1995).
RN   [4]
RP   SEQUENCE OF 40-53.
RX   MEDLINE=87204112; PubMed=3033641;
RA   Takada Y., Strominger J.L., Hemler M.E.;
RT   "The very late antigen family of heterodimers is part of a
RT   superfamily of molecules involved in adhesion and embryogenesis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:3239-3243(1987).
RN   [5]
RP   SEQUENCE OF 598-607, AND PROCESSING.
RX   MEDLINE=92112902; PubMed=1730718;
RA   Teixido J., Parker C.M., Kassner P.D., Hemler M.E.;
RT   "Functional and structural analysis of VLA-4 integrin alpha 4 subunit
RT   cleavage.";
RL   J. Biol. Chem. 267:1786-1791(1992).
CC   -!- FUNCTION: INTEGRINS ALPHA-4/BETA-1 (VLA-4) AND ALPHA-4/BETA-7 ARE
CC       RECEPTORS FOR FIBRONECTIN. THEY RECOGNIZE ONE OR MORE DOMAINS
CC       WITHIN THE ALTERNATIVELY SPLICED CS-1 AND CS-5 REGIONS OF
CC       FIBRONECTIN. THEY ARE ALSO RECEPTORS FOR VCAM1. INTEGRIN ALPHA-
CC       4/BETA-1 RECOGNIZES THE SEQUENCE Q-I-D-S IN VCAM1. INTEGRIN
CC       ALPHA-4/BETA-7 IS ALSO A RECEPTOR FOR MADCAM1. IT RECOGNIZES THE
CC       SEQUENCE L-D-T IN MADCAM1. ON ACTIVATED ENDOTHELIAL CELLS INTEGRIN
CC       VLA-4 TRIGGERS HOMOTYPIC AGGREGATION FOR MOST VLA-4-POSITIVE
CC       LEUKOCYTE CELL LINES. IT MAY ALSO PARTICIPATE IN CYTOLYTIC T-CELL
CC       INTERACTIONS WITH TARGET CELLS.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC       SUBUNIT CAN SOMETIMES BE CLEAVED INTO TWO NON-COVALENTLY
CC       ASSOCIATED FRAGMENTS. ALPHA-4 ASSOCIATES WITH EITHER BETA-1 OR
CC       BETA-7.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD49d entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49d.htm".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X16983; CAA34852.1; -.
DR   EMBL; L12002; AAB59613.1; -.
DR   PIR; S06046; S06046.
DR   PIR; D28018; D28018.
DR   HSSP; P11215; 1A8X.
DR   MIM; 192975; -.
DR   InterPro; IPR000413; Integrin_alpha.
DR   Pfam; PF01839; FG-GAP; 5.
DR   Pfam; PF00357; integrin_A; 1.
DR   PRINTS; PR01185; INTEGRINA.
DR   SMART; SM00191; Int_alpha; 5.
DR   PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW   Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW   Signal; Repeat; Polymorphism; Calcium.
FT   SIGNAL        1     39
FT   CHAIN        40   1038       INTEGRIN ALPHA-4.
FT   DOMAIN       41    983       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    984   1007       POTENTIAL.
FT   DOMAIN     1008   1038       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       55    117       FG-GAP 1.
FT   REPEAT        ?      ?       FG-GAP 2.
FT   REPEAT      194    253       FG-GAP 3.
FT   REPEAT      254    306       FG-GAP 4.
FT   REPEAT      309    368       FG-GAP 5.
FT   REPEAT      371    430       FG-GAP 6.
FT   REPEAT      433    485       FG-GAP 7.
FT   CA_BIND     320    328       POTENTIAL.
FT   CA_BIND     383    391       POTENTIAL.
FT   CA_BIND     445    453       POTENTIAL.
FT   SITE        597    598       CLEAVAGE.
FT   SITE       1009   1013       GFFKR MOTIF.
FT   DISULFID     97    107       BY SIMILARITY.
FT   DISULFID    150    171       BY SIMILARITY.
FT   DISULFID    189    204       BY SIMILARITY.
FT   DISULFID    492    501       BY SIMILARITY.
FT   DISULFID    507    563       BY SIMILARITY.
FT   DISULFID    628    633       BY SIMILARITY.
FT   DISULFID    704    717       BY SIMILARITY.
FT   DISULFID    858    896       BY SIMILARITY.
FT   DISULFID    903    908       BY SIMILARITY.
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    235    235       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    486    486       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    524    524       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    544    544       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    632    632       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    812    812       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     884    884       R -> Q.
FT                                /FTId=VAR_003978.
FT   MUTAGEN     596    596       K->Q: ABOLISHES ALMOST COMPLETELY
FT                                CLEAVAGE.
FT   MUTAGEN     597    597       R->L: ABOLISHES COMPLETELY CLEAVAGE.
SQ   SEQUENCE   1038 AA;  115332 MW;  ED7D9075FAEB8996 CRC64;

alignment_scores:
           Quality:  119.00           Length:    482
             Ratio:  0.511              Gaps:     25
  Percent Similarity:  48.340   Percent Identity: 23.651

alignment_block:
 US-09-303-518D-463/rev x ITA4_HUMAN   ..

 Align seg 1/1  to: ITA4_HUMAN  from: 1  to: 1038

    1352 ACAGGCTTAGCCTTAGGTATACCGCCCCCCGATAATTCTTGAATATCGAG 1303
         ||||||   :::   |||:::||||||      ::::::   :::|||::
     168 ThrGlyGlyCysTyrGlyValProProAspLeuArgThrGluLeuSerLy 184

    1302 CTTCGTATCA......TATTTCACGTGCTTCTCAAAATTCGGAAACCCTT 1259
         :   ::::::       |||             :::|||||||||:::   :
     184 sArgIleAlaProCysTyrGlnAspTyrValLysLysPheGlyGluAsnP 201

    1258 TACCGTCA.......................................... 1251
         ::   |||
     201 heAlaSerCysGlnAlaGlyIleSerSerPheTyrThrLysAspLeuIle 217

    1250 ...AACGGTACGCCTGTCTTCGGGTGGCGTTGGTCTGCCAGTTTGACATT 1204
                |||:::|||      ::::::   |||::::::||||||
     218 ValMetGlyAlaProGlySerSerTyr...TrpThrGlySerLeu.PheV 233

    1203 TTTGCCGTTTGACGGCGGCACGGTTGAGGAGGTGATGTTTTCTTTGCCGT 1154
               ::::::|||:::    :::    :::        :::||||||    :::
     233 alTyrAsnIleThrThrAsnLysTyrLys......AlaPheLeuAspLys 247

    1153 AACGCTGCTCCAAGTTTGAACGGATATTTCGGGAATGGTAAGGGGACGGG 1104
         :::           ::::::::::::|||:::|||    :::    |||:::|||
     248 GlnAsnGlnValLysPheGlySerTyrLeuGlyTyrSerValGlyAlaGl 264

    1103 TATTTGGCGTATGCCGCATCGGCAAAATTGTCGCTGACGGCGGATTTCCC 1054
```

```
           |  :::|||      |||      ::: |||:::   |||:::
  264 yHisPheArgSerGlnHisThrThrGlu.ValValGlyGlyAlaProGln 280

 1053 TTTCGGCAATGCGATCGCGCCCATCTGCGACCGCTTGACAGGATGTGCCG 1004
         :::|||      :::|||:::::     :::|||:::
  281 HisGluGlnIleGlyLysAlaTyrIlePheSerIleAspGluLysGluLe 297

 1003 TGATGCCGCCCAAGCCGTATTTTCCCCGGACAGCTCCAATCCCTTTGATG 954
       :::       :::   :::   :::        ::::::  |||   :
  297 uAsnIleLeuHisGluMetLysGlyLysLysLeuGlySerTyrPheGlyA 314

  953 GGGATGGCTGCCATAAAGATATT...GCTGACGGCTTCTATGCCTTGTGC 907
      :::::  |||      |||:::   |||||||||||        |||
  314 laSerValCysAlaValAspLeuAsnAlaAspGlyPheSerAspLeuLeu 330

  906 GGCATTGGGGTTTTGGACTGCCCAATCGCGGATGGCTGCTGCGGCATAGT 857
             :::   ::::::       |||   :::
  331 ValGlyAlaProMetGlnSerThrIleArgGlu.............. 341

  856 CTTTGAGTTGCGCCATATCTGCCAAATCGTTGATGCGCGCCATCTTGTTT 807
          |||   |||   :::     |||   ::::::       |||:
  342 ....GluGlyArgValPheValTyrIleAsnSerGlySerGlyAlaValM 357

  806 TCGGTGGAAAGCAGACCCAAGCCGTG...CATGACAGCAATGTTTGAGCC 760
      ::::::::    :::|||:::    |||      |||:::   :::
  357 etAsnAlaMetGluThrAsnLeuValGlySerAspLysTyrAlaAlaArg 373

  759 TTCGCTTATACCCTGCACGGCATCGCCTGCGCCGACAATTTCTCCTGCCG 710
      |||:::    ::::::    :::::::::          ||||||
  374 PheGlyGluSerIleValAsnLeuGlyAspIleAspAsn........... 386

  709 CGCCGATGATGTTTTGACGATATCTGCAGTGCCGTTGAAGGCTTCGGCG 660
              |||   |||:::|||
  387 .......AspGlyPheGluAsp........................... 391

  659 GCATTGCCCGATCTGTCCAGCTCGGGGCTGTATCGGGTGGCGCGTTTGAA 610
                          |||   :::||||||   :::      ::::::
  392 .............ValAlaIleGlyAlaProGlnGluAspAspLeuGln 403

  609 TCCGTCGCCTACTCCTTGCGTCAGCATAGCGCCGGCATTGTGGAAACG.. 562
      ::::::  |||      :::   ::::::    ||||||   :::|||
  404 GlyAlaIleTyrIleTyrAsnGlyArgAlaAspGlyIleSerSerThrPh 420

  561 .........................GTCGGCAAGCCGTTGTCC......G 543
                               ::::::|||   ||||||    |
  420 eSerGlnArgIleGluGlyLeuGlnIleSerLysSerLeuSerMetPheG 437

  542 GTGCTGCGGTTGTCGGTCAGGTTGAGGCGGATATTTGGGCAACGCCTTT 493
       ||   ::::::   ||||||::::::||||||         :::|||::::::
  437 lyGlnSerIleSerGlyGlnIleAspAlaAsp......AsnAsnGlyTyr 451

  492 TATGTCGTAGCTGTATATATCCCTCGCGCCTTTGGGAGCGGGATAGCCGC 443
               |||||||||           ||||||   |||     |||::
  452 ValAspValAlaVal.........GlyAlaPheArgSerAspSerAlaVa 465

  442 CGCCCTG......TGGCCCGTCATAGCCGTCGGCGGGATGGTGTTCGTAT 399
       :   |||         ||||||:::   |||   :::        :::|
  465 lLeuLeuArgThrArgProValValIleValAspAlaSer......LeuS 480

  398 CCGTCCCAATGGATGCGGTAAAGGCTGAATCCGTCAACGGGACTACC... 352
       ||   |||                   ||||||||||||   |||
  480 erHisPro..................GluSerValAsnArgThrLysPhe 490

  351 GGCTTCGTCAGAATCGGAATGTGAGGCATGGTTGTCGAAGGGCGAATGGA 302
               |||:::   |||      ::::::   ::::::     :::
  491 AspCysValGluAsnGlyTrpProSerValCysIleAspLeuThrLeuCy 507

  301 ATTTGTGCCCGTGATCGGAAAAGCGGACAATGTAGCCGATATTTCCTTCA 252
        |||                                      |||:::::
  507 sPhe.....................................SerTyrL 511

  251 ACGGCCGCCTGTTGAATCATCAGGTGGCCCAACTGATGGCTTTGTATGTT 202
       ::|||:::               :::|||            ::::::|||:::

  511 ysGlyLys............GluValProGlyTyrIleValLeuPheTyr 523

  201 TCCCAATCCGATATGGCCGTTGCGCTNGGCAAGCTCCCCCCTGCTGCCGA 152
      ::::::|||  :::            |||   |||:::|||        |||
  524 AsnMetSer.LeuAspValAsnArgLysAlaGluSer.........ProP 537

  151 ATAGGTGGTATTTCCCGTCGGGTTCGAAATGCTGACGGTCGAGAACCTGC 102
        |||:::||||||    |||                 ::::::    :::
  537 roArgPheTyrPheSerSer...............AsnGlyThrSerAsp 548

  101 CGGATAAAGGGATCGTTTGCCAAATCTGAGGCGTGTGCATGCATCGGCAG 52
         |||    ||||||:::         |||:::       |||
  549 ValIleThrGlySerIleGlnValSerSerArgGluAlaAsnCysArgTh 565

   51 GCACACTGCCAGTATGGACAGA...ATAAGGGATATTTTG 15
        |||    |||    |||      :::    :::||||||||||||
  565 rHisGlnAlaPheMetArgLysAspValArgAspIleLeu 578
```

seq_name: SwissProt_40:MAP4_BOVIN

```
seq_documentation_block:
ID   MAP4_BOVIN     STANDARD;      PRT;  1072 AA.
AC   P36225;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 4 (MAP 4) (Microtubule-associated
DE   protein-U) (MAP-U).
GN   MAP4.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90338002; PubMed=2380192;
RA   Aizawa H., Emori Y., Murofushi H., Kawasaki H., Sakai H.,
RA   Suzuki K.;
RT   "Molecular cloning of a ubiquitously distributed
RT   microtubule-associated protein with Mr 190,000.";
RL   J. Biol. Chem. 265:13849-13855(1990).
RN   [2]
RP   DOMAINS.
RX   MEDLINE=91236765; PubMed=2033072;
RA   Aizawa H., Emori Y., Mori A., Murofushi H., Sakai H., Suzuki K.;
RT   "Functional analyses of the domain structure of
RT   microtubule-associated protein-4 (MAP-U).";
RL   J. Biol. Chem. 266:9841-9846(1991).
CC   -!- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES
CC       MICROTUBULE ASSEMBLY.
CC   -!- TISSUE SPECIFICITY: IS DISTRIBUTED UBIQUITOUSLY AMONG ALL TISSUES
CC       BUT AMOUNTS ARE LOWER IN CEREBELLUM AND LIVER.
CC   -!- PTM: PHOSPHORYLATION OF THE PRO-RICH REGION IN THE C-TERMINUS
CC       NEGATIVELY REGULATES MAP-4 ACTIVITY TO PROMOTE MICROTUBULE
CC       ASSEMBLY.
CC   -!- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D90149; BAA14179.1; -.
DR   PIR; A37127; A37127.
DR   InterPro; IPR001084; Tubulin-bind.
DR   Pfam; PF00418; tubulin-binding; 3.
DR   PROSITE; PS00229; TAU_MAP; 3.
KW   Microtubules; Repeat; Phosphorylation.
FT   DOMAIN      244    530       19 X 14 AA TANDEM REPEATS.
```

```
FT   REPEAT      244    257       1.
FT   REPEAT      258    271       2.
FT   REPEAT      272    285       3.
FT   REPEAT      286    299       4.
FT   REPEAT      300    313       5.
FT   REPEAT      314    327       6.
FT   REPEAT      328    341       7.
FT   REPEAT      342    355       8.
FT   REPEAT      384    391       9 (INCOMPLETE).
FT   REPEAT      392    405      10.
FT   REPEAT      406    417      11.
FT   REPEAT      418    431      12.
FT   REPEAT      432    445      13.
FT   REPEAT      446    460      14.
FT   REPEAT      461    474      15.
FT   REPEAT      475    488      16.
FT   REPEAT      489    502      17.
FT   REPEAT      503    516      18.
FT   REPEAT      517    530      19.
FT   REPEAT      907    937      TAU/MAP MOTIF.
FT   REPEAT      938    968      TAU/MAP MOTIF.
FT   REPEAT      969   1000      TAU/MAP MOTIF.
SQ   SEQUENCE   1072 AA;  111914 MW;  E8C17A730989F0D2 CRC64;

alignment_scores:
           Quality:  118.50             Length:     560
             Ratio:  0.459                Gaps:      31
  Percent Similarity:  46.071    Percent Identity:  22.321

alignment_block:
 US-09-303-518D-463 x MAP4_BOVIN   ..

 Align seg 1/1   to: MAP4_BOVIN   from: 1   to: 1072

     308 CGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGT 357
          :::||||||::          ::::::|||     |||:::   |||::
     600 LysProSerSerGlnThrSerGluLeuProSerGluThrSerGlyValAl 616

     358 CCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACA 407
          :          ::::::      |||:::|||::::::   :::      |
     616 aLysProGluGluGlyProProThrGlySerValSerGlyAsnAspIleT 633

     408 CCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTC 457
          ||  ||||||                      :::          |||
     633 hrAlaProPro.................AsnLysGluLeuProProSer 643

     458 CCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCG........ 499
          |||:::   :::            |||||||||     |||
     644 ProGluLysLysThrLysProLeuAlaThrThrGlnProAlaLysThrSe 660

     500 ..............................TTGCCCAAAATATCCGCCT 518
                                       |||||||||    :::|||
     660 rThrSerLysAlaLysThrGlnProThrSerLeuProLysGlnThrAla. 676

     519 CAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCG......... 559
                  |||||||||    :::    :::|||       |||
     677 ........ProThrThrLeuGlyGlySerAsnLysLysProMetSerLeu 690

     560 ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGA 609
          :::    :::::::::||||||          ::::::       |||:::
     691 AlaSerGlySerValProAlaAlaProProLysArgProAlaAlaAlaTh 707

     610 TTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
          |||:::       |||:::|||   ||||||            :::
     707 rSerArg......ProSerThrLeuProSer.........LysAspThrL 719

     660 CGCCGAAGCCTTCAACGGCACTGCAGATA.................TCG 691
            |||||||||    :::       :::|||
     719 ysProLysProValAlaGluAlaLysIleProGluLysArgValSerPro 735

     692 TCAAAAACATCATCGGCGCGGCAGGAGAAA.........TTGTCGGCGCA 732
          ||||||    :::||||||         |||          |||:::||
     736 SerLysProAlaSerAlaProAlaValLysProGlySerLysSerThrGl 752

     733 GGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCA 782
          ||||:::|||                ||||||              :::|
     752 nAlaValPro..............LysAla...........ProAlaT 760

     783 CGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATT 832
          |||||    ::::::    |||    :::|||        :::    ||||||
     760 hrAlaThrLeuAlaSerProGlySerThrSerArgAsnLeuSerThr... 775

     833 TGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGAT 882
                                 :::|||::::::|||:::|||||
     776 ......................ProLeuProLysArgProThrAlaIl 784

     883 TGGCAGTCCAAAACCCCAATGCCGCACAAG......GCATAGAAGCCGT 926
          |        :::       |||         |||       |||   |||    :
     784 eLysThrGluGlyLysProAlaGluIleLysLysMetAlaThrLysSerA 801

     927 CAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCC 976
          ::         :::|||    :::|||:::        :::        ::::::::
     801 laProAlaAspLeuSerArgProLysSerThrThrThrSerSerValLys 817

     977 GGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCG 1026
            ::::::|||:::    :::::::::
     818 LysSerThrThrValProGlyThrAla...................... 826

    1027 CAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAA 1076
                                             ||||||::::::::::
     827 ................................ProProAlaGlyAla. 831

    1077 TTTTGCCGATGCGGCATACGCCAAATACC.........CGTCCCCTTACC 1117
                |||    |||       |||::::::        ||||||
     832 .....ProSerArgAlaArgProThrAlaThrProProArgProSerGly 846

    1118 ATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAAC 1167
               |||    ::::::    :::             ::::::||||||    ||
     847 ThrProProValAspLysLys.........ProThrAlaAlaLysProTh 860

    1168 ATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGC 1217
          ||||                          |||    :::|||   |||::::::
     860 rSerSerAlaProArgLeuGlyArgValAlaAla..AsnAlaSerAlaPr 876

    1218 AGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTC 1267
          |||    ::::::    ::::::    |||                 |||
     876 oAspLeuLysAsnValArgSerLysVal..............GlySerT 888

    1268 CGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAA 1317
             :::      ||||||              :::       ::::::::
     888 hrGluAsnIleLysHisGlnProGlyGlyGlyArgAlaLysValGluLys 904

    1318 TTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCG..... 1362
             :::    ::::::    |||       ||||||       :::|||
     905 LysThrGluAlaAlaAlaProAlaArgLysProGluProAsnAlaValTh 921

    1363 .............................AAACCG.........AGAT 1372
                                        ||||||           :::
     921 rLysAlaAlaGlyProIleGlyAsnAlaGlnLysProProThrGlyLysV 938

    1373 GGGAGGTTGATAGGAAG.........CTTAATAAATTGACAACTCGTGAG 1413
             :::::::::::|||          ::::::|||:::::::::::::
     938 alGlnIleGlnAsnLysLysValAspIleSerLysValSerSerLysCys 954

    1414 CAGGTGGAGAAAAATGTTCAG.................GAAACGAGAAG 1445
               :::    |||::::::                  :::::::::
     955 GlySerLysAlaAsnIleLysHisLysProGlyGlyGlyAspValLysIl 971

    1446 AAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAA 1495
          :::||||||:::    :::||||||        ||||||
     971 eGluSerGlnLysLeuAsnPheLysGluLysAlaGlnAla.......... 984

    1496 ATAAAACAGGG...TTAGAT...TTTAATCATTTT...ATAGGTGGTGAT 1536
          |||:::|||    ||||||     :::|||:::       ||||||
     985 ..LysValGlySerLeuAspAsnValGlyHisLeuProAlaGlyGlyAla 1000
```

```
1537 ATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGTGA 1586
     ::::::   :::|||        ||||||
1001 ValLysThrGluGly.........GlyGly................... 1007

1587 TGTACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTTTATC 1636
                       :::::||||||         |||
1008 ..................SerGluAlaProProCysProGlyPro...P 1017

1637 AAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTGAAAACG 1686
       |||      |||:::       |||:::::::::    |||:::   :::
1017 roAlaGlyGluGluLeuAlaIleProGluAlaAlaProGluAlaGlyAla 1033

1687 AAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAGATTG 1736
           :::::::::   ::::::   ||||||   :::
1034 ProAlaSerAlaSerGlyLeuSerGlyHisThrThrLeuAlaGlyGlyGl 1050

1737 GGATGAGGCTAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAA 1786
       |||:::   :::                 :::::::::   :::|||:::|
1050 yAspGlnArgGlu..............AlaGlnThrLeuAspSerGlnI 1062

1787 TAATGCTTAAGGATAATAAATGGCAGGGT 1815
     ||:::   :::   :::||||||   |||
1062 leGln...GluThrSerLysTrpLeuGly 1070

seq_name: SwissProt_40:P2P_LACLC

seq_documentation_block:
ID   P2P_LACLC      STANDARD;      PRT;  1902 AA.
AC   P15293;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
DE   associated serine proteinase) (LP151).
GN   PRT.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG   Plasmid pLP763.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCDO 763;
RX   MEDLINE=89313288; PubMed=2501630;
RA   Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;
RT   "Molecular characterization of a cell wall-associated proteinase gene
RT   from Streptococcus lactis NCDO763.";
RL   Mol. Microbiol. 3:359-369(1989).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC       specificity, although some subsite preference have been noted,
CC       e.g. large hydrophobic residues in the P1 and P4 positions, and
CC       Pro in the P2 position. Best known for its action on caseins,
CC       although it has been shown to hydrolyse hemoglobin and oxidized
CC       insulin B-chain.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X14130; CAA32350.1; -.
DR   PIR; S06997; S06997.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.019; -.
```

```
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR003137; PA.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF02225; PA; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW   Transmembrane.
FT   SIGNAL        1     33       POTENTIAL.
FT   PROPEP       34    187       POTENTIAL.
FT   CHAIN       188   1902       PII-TYPE PROTEINASE.
FT   DOMAIN      188   1876       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1877   1895       MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN     1896   1902       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE    217    217       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    281    281       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    620    620       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DOMAIN     1867   1872       CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                                PROTEINS.
SQ   SEQUENCE   1902 AA;  200139 MW;  4B8D8B844D88CDF7 CRC64;

alignment_scores:
           Quality:  118.50           Length:    672
             Ratio:  0.386              Gaps:     33
 Percent Similarity:  45.685   Percent Identity: 19.345

alignment_block:
 US-09-303-518D-463 x P2P_LACLC   ..

 Align seg 1/1  to: P2P_LACLC  from: 1  to: 1902

 271 ATTGTCCGCTTTTCCGAT.........CACGGGCACAAATTCCATTCG.. 309
     :::   :::|||:::|||          ||||||:::   |||:::|||
 242 ValGluLysPheThrAspThrAlaLysHisGlyArgTyrPheAsnSerLy 258

 310 ....CCCTTCGAC...AACCATGCCTCACATTCCGAT......TCTGACG 346
         |||:::        |||:::|||:::::::::|||       :::|||:
 258 sValProTyrGlyPheAsnTyrAlaAspAsnAsnAspThrIleThrAspA 275

 347 AAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGAC 396
     :::::          ||||||
 275 spThr.........ValAsp............................. 278

 397 GGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGG 446
         |||:::|||                    |||   :::   :::||
 279 ......GluGlnHis..................GlyMetHisValAlaGl 287

 447 CTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAG 496
     |      :::    :::|||:::   |||      :::   ::::::   |
 287 yIleIleGlyAlaAsnGlyThrGlyAspAspProAlaLysSerValValG 304

 497 GCGTTGCCCAAAATATCCGCCTC.....................AACCTG 525
     ||||||||   :::   :::|||                      |||
 304 lyValAlaProGluAlaGlnLeuLeuAlaMetLysValPheThrAsnSer 320

 526 ACCGACAACCGCAGCACCGGA..............CAACGGCTTGCCGA 560
       :::   :::||||||                 :::   :::   ||
 321 AspThrSerAlaThrThrGlySerAlaThrLeuValSerAlaIleGluAs 337

 561 CCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGAT 610
     |          ::::::   :::|||:::::::::::::|||:::
 337 pSerAlaLysIleGlyAlaAspValLeuAsnMetSerLeuGlySerAspS 354

 611 TCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCC 660
       :::   |||      :::|||||||||            ||||||
 354 erGlyAsnGlnThrLeuGluAspProGluLeuAlaAlaValGlnAsnAla 370
```

```
 661 GCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGC 710
      |||        :::|||||||||  ::::::               ::
 371 AsnGlu......SerGlyThrAlaAlaValIle...............Se 380

 711 GGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATA....... 753
     :||||||:::       :::|||:::|||::::::|||:::
 380 rAlaGlyAsnSerGlyThrSerGlySerAlaThrGluGlyValAsnLysA 397

 754 ...............................................AGC 756
                                                    |||
 397 spTyrTyrGlyLeuGlnAspAsnGluMetValGlyThrProGlyThrSer 413

 757 GAAGGCTCAAAC...................................... 768
     :::|||::::::
 414 ArgGlyAlaThrThrValAlaSerAlaGluAsnThrAspValIleThrGl 430

 769 ....ATTGCTGTCATGCACGGCTTGGGTCTG...CTTTCCACCGAAAACA 811
         :::::::::     |||  ||||||  |||:::  |||:::
 430 nAlaValThrIleThrAspGlyThrGlyLeuGlnLeuGlyProGluThrI 447

 812 AGATGGCGCGCATCAACGATTTGGCA........................ 837
        :::          ||||||:::::
 447 leGlnLeuSerSerAsnAspPheThrGlySerPheAspGlnLysLysPhe 463

 838 ..............................GATATGGCGCAACTCAAAGA 857
                                   :::  :::::::::   ||
 464 TyrValValLysAspAlaSerGlyAsnLeuSerLysGlyLysValAlaAs 480

 858 CTATGCCGCAGCAGCCATCCGCGATTGGGCAGTC................ 891
     ||||:::|||   |||          |||:::
 480 pTyrThrAlaAspAlaLysGlyLysIleAlaIleValLysArgGlyGluL 497

 892 .........................CAAAACCCCAATGCCGCACAAGGC 915
                                  :::     :::|||  |||
 497 euThrPheAlaAspLysGlnLysTyrAlaGlnAlaAlaGlyAlaAlaGly 513

 916 ATAGAAGCCGTCAGCAAT..............................AT 935
     :::     |||:::|||
 514 LeuIleIleValAsnAsnAspGlyThrAlaThrProValThrSerMetAl 530

 936 CTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAAT 985
       :::   :::::::::|||     |||::::::::::|||   |||:::
 530 aLeuThrThrThrPheProThrPheGlyLeuSerSerValThrGlyGlnL 547

 986 ACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGC 1035
          :::      ::::::|||||||||          :::::::|||
 547 ysLeuValAspTrpValAlaAlaHisPro......AspAspSerLeuGly 561

1036 .........GCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAA 1076
                  |||:::::|||       ::::::::  ::::::|||::
 562 ValLysIleAlaLeuThrLeuValProAsnGlnLysTyrThrGluAspLy 578

1077 TTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAA 1126
     ::::::|||         :::::::|||   :::|||      :::   :
 578 sMetSerAsp......PheThrSerTyr...GlyProValSerAsnLeuS 592

1127 ATATCCGTTCAAACTTG..............................GAG 1146
     ::::::::  ::::::                               :::
 592 erPheLysProAspIleThrAlaProGlyGlyAsnIleTrpSerThrGln 608

1147 CAGCGTTACGGCAAAGAAAACATC......ACCTCCTCAACCGTGCCGCC 1190
     ::::::   |||       |||:::      ||||||   :::    |||
 609 AsnAsnAsnGlyTyrThrAsnMetSerGlyThrSerMetAlaSerProPh 625

1191 GTCAAACGGCAAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAG 1240
              |||:::::::::   |||   :::        :::  |||:::
 625 eIleAlaGlySerGlnAlaLeuLeuLysGlnAlaLeuAsnAsnLysAsnA 642

1241 GCGTACCGTTT....................GACGGTAAAGGGTTTCCG 1269
     ::     ||||||                       |||   ::::::
 642 sn...ProPheTyrAlaTyrTyrLysGlnLeuLysGlyThrAlaLeuThr 657

1270 AATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATT 1319
     :::|||   |||    |||:::   :::|||             |||::::::::
 658 AspPheLeuLysThrValGluMetAsnThrAlaGlnProIleAsnAspIl 674

1320 ATCG................................GGGGGCGGTATAC 1336
     ::::                                |||:::|||:::
 674 eAsnTyrAsnAsnValIleValSerProArgArgGlnGlyAlaGlyLeuV 691

1337 CTAAGGCTAAGCCTGTGTTTGATGCG.........AAACCGAGATGGGAG 1377
          :::|||   ::::::||||||          :::|||
 691 alAspValLysAlaAlaIleAspAlaLeuGluLysAsnProSerThrVal 707

1378 GTTGATAGG.................AAGCTTAATAAATTGACAACTCG 1409
     |||    :::                :::|||:::    :::|||:::
 708 ValAlaGluAsnGlyTyrProAlaValGluLeuLysAspPheThrSerTh 724

1410 TGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTA 1459
        :::::::::   |||    :::      |||:::|||   ::::::::
 724 rAspLysThrPheLys...LeuThrPheThrAsnArgThrThrHisGluL 740

1460 GTCAGTTTAAAGCCCATGCGCAACGAGAA..................... 1488
           ::::::      ::::::   :::
 740 euThrTyrGlnMetAspSerAsnThrAspThrAsnAlaValTyrThrSer 756

1489 ...TGGGAAAATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGA 1535
           :::    ::::::|||:::   ::::::      |||   |||
 757 AlaThrAspProAsnSerGlyValLeuTyrAspLysLysIleAspGlyAl 773

1536 TATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGTG 1585
        |||   |||:::            :::   ::::::      |||
 773 aAlaIleLysAlaGlySer.........AsnIleThrValProAlaGlyL 787

1586 ATGTACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTTTAT 1635
       :::          |||:::   |||   :::       |||
 787 ysThrAlaGlnIleGluPheThrLeuSerLeuProLysSerPheAspGln 803

1636 CAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTGAAAAC 1685
     |||      ||||||                |||      :::   |||
 804 GlnGlnPheValGlu...............GlyPheLeuAsnPheLysGl 815

1686 GAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAGATT 1735
       :::    |||:::::::::          :::      |||     ||||
 815 ySerAspGlySerArgLeuAsnLeuProTyrMetGlyPhePheGlyAspT 832

1736 GGGATGAGGCTAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAGT... 1782
     ||:::::::::::|||      ::::::::::::           |||
 832 rpAsnAspGlyLysIleValAspSerLeuAsnGlyIleThrTyrSerPro 848

1783 .................AGAATAATGCTTAAGGATAATAAATGGCAGGG 1814
                          :::   |||    ||||||      ||
 849 AlaGlyGlyAsnPheGlyThrValProLeuLeuThrAsnLysAsnThrGl 865

1815 TACAAGTAAATCGGGT 1830
      ||||:::      |||
 865 yThrGlnTyrTyrGly 870
```

seq_name: SwissProt_40:CCAA_RAT

seq_documentation_block:

```
ID   CCAA_RAT       STANDARD;      PRT;   2212 AA.
AC   P54282; Q01541;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium
DE   channel, L type, alpha-1 polypeptide, isoform 4) (Brain calcium
DE   channel I) (BI) (RAT brain class A) (RBA-I).
GN   CACNA1A OR CACNL1A4 OR CACH4 OR CACN3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=91288516; PubMed=1648226;
RA   Starr T.V.B., Prystay W., Snutch T.P.;
RT   "Primary structure of a calcium channel that is highly expressed in
RT   the rat cerebellum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5621-5625(1991).
RN   [2]
RP   SEQUENCE OF 1435-1667 FROM N.A. (CLONE RBA-65).
RX   MEDLINE=90239020; PubMed=1692134;
RA   Snutch T.P., Leonard J.P., Gilbert M.M., Lester H.A., Davidson N.;
RT   "Rat brain expresses a heterogeneous family of calcium channels.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:3391-3395(1990).
RN   [3]
RP   SEQUENCE OF 1354-1659 FROM N.A. (CLONE RKC8).
RC   TISSUE=Kidney;
RX   MEDLINE=93066265; PubMed=1279681;
RA   Yu A.S.L., Hebert S.C., Brenner B.M., Lytton J.;
RT   "Molecular characterization and nephron distribution of a family of
RT   transcripts encoding the pore-forming subunit of Ca2+ channels in the
RT   kidney.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10494-10498(1992).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A
CC       GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM
CC       CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND
CC       ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-
CC       IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO
CC       DIHYDROPYRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-
CC       GVIA).
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: IN THE BRAIN, AT LEAST FOUR ISOFORMS ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC; PURKINJE CELLS CONTAIN
CC       PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM
CC       CURRENT IN CEREBELLAR GRANULE CELLS. ALSO FOUND IN HEART, IN
CC       KIDNEY DISTAL CONVOLUTED TUBULE (DCT), AND IN PITUITARY.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M64373; AAA40806.1; -.
DR   EMBL; M99222; AAA40896.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
DR   SMART; SM00384; AT_hook; 1.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation; Alternative splicing.
FT   REPEAT       87    365       I.
FT   REPEAT      475    719       II.
FT   REPEAT     1182   1465       III.
FT   REPEAT     1502   1765       IV.
FT   DOMAIN        1    100       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    101    119       S1 OF REPEAT I (POTENTIAL).
FT   DOMAIN      120    138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    139    156       S2 OF REPEAT I (POTENTIAL).
FT   DOMAIN      157    168       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    169    184       S3 OF REPEAT I (POTENTIAL).
FT   DOMAIN      185    192       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    193    211       S4 OF REPEAT I (POTENTIAL).
FT   DOMAIN      212    230       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    231    250       S5 OF REPEAT I (POTENTIAL).
FT   DOMAIN      251    337       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    338    362       S6 OF REPEAT I (POTENTIAL).
FT   DOMAIN      363    489       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    490    509       S1 OF REPEAT II (POTENTIAL).
FT   DOMAIN      510    523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    524    543       S2 OF REPEAT II (POTENTIAL).
FT   DOMAIN      544    551       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    552    570       S3 OF REPEAT II (POTENTIAL).
FT   DOMAIN      571    580       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    581    599       S4 OF REPEAT II (POTENTIAL).
FT   DOMAIN      600    618       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    619    638       S5 OF REPEAT II (POTENTIAL).
FT   DOMAIN      639    691       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    692    716       S6 OF REPEAT II (POTENTIAL).
FT   DOMAIN      717   1190       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1191   1214       S1 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1215   1231       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1232   1251       S2 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1252   1258       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1259   1282       S3 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1283   1293       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1294   1311       S4 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1312   1330       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1331   1350       S5 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1351   1437       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1438   1462       S6 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1463   1518       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1519   1537       S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1538   1551       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1552   1573       S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1574   1580       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1581   1600       S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1601   1607       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1608   1626       S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1627   1645       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1646   1665       S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1666   1737       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1738   1763       S6 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1764   2212       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       13     20       POLY-GLY.
FT   DOMAIN      729    734       POLY-GLU.
FT   DOMAIN     1155   1158       POLY-GLU.
FT   DOMAIN     2162   2171       POLY-HIS.
FT   DOMAIN      385    402       BINDING TO THE BETA SUBUNIT (BY
FT                                SIMILARITY).
FT   SITE        320    320       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE        670    670       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1411   1411       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1707   1707       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   MOD_RES    1773   1773       PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND    1791   1802       BY SIMILARITY.
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1607   1607       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   VARSPLIC    1602   1602       G -> GNP (IN ISOFORM RKC8).
SQ   SEQUENCE    2212 AA;  251525 MW;  DFBD7A0F553AFA52 CRC64;

alignment_scores:
            Quality:  118.50            Length:    461
              Ratio:  0.620               Gaps:     27
 Percent Similarity:  41.432   Percent Identity: 23.861

alignment_block:
 US-09-303-518D-463 x CCAA_RAT   ..

 Align seg 1/1   to: CCAA_RAT   from: 1   to: 2212

     183 CGGCCATATCGGATTGGGAAACATACA.........AAGCCATCAGTTGG 223
          ||||||   |||          |||        :::|||   |||
     810 ArgProLeuArgProAspValLysThrHisLeuAspArgProLeuValVa 826

     224 GCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATT 273
               |||:::::::::   ::::::::   :::|||
     826 lAspProGlnGluAsnArgAsnAsnAsnThrAsnLys............. 838

     274 GTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCA 323
                       |||||||||     |||   |||||||||||
     839 ..............SerArgAla......ProGluAlaLeuArgGlnThr 848

     324 TGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCA 373
         :::
     849 Ala............................................... 849

     374 GCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGC 423
                    :::|||             |||   |||:::   ||||||
     850 ..........ArgProArgGluSerAlaArgAspProAspAlaArgArgA 863

     424 TA...............TGACGGGCCACAGGGCGG............... 443
                             ||||||   ||||||
     863 laTrpProSerSerProGluArgAlaProGlyArgGluGlyProTyrGly 879

     444 CGGCTATCCCGCTCCCAAAGGCGCGAG....................... 470
         |||   |||:::   |||:::||||||
     880 ArgGluSerGluProGlnGlnArgGluHisAlaProProArgGluHisVa 896

     471 .................................GGATATATACAGCTAC 486
                                        |||          |
     896 lProTrpAspAlaAspProGluArgAlaLysAlaGlyAspAlaProArgA 913

     487 GACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCG 536
          |||||   :::|||               |||      :::   :::|||
     913 rgHisThrHisArg...............ProValAlaGluGlyGluPro 924

     537 CAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGC 586
         ::::::::   :::         ||||||      :::
     925 ArgArgHisArgAlaArgArgArgProGlyAspGlu............Pr 937

     587 TGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCC. 635
          |||   |||   :::||||||   :::   |||         :::|||
     937 oAspAspArgProGluArgArgProArgProArgAspAlaThrArgProA 954

     636 ..CGAGCT............GGACAGATCGGGCAATGCCGCCGAAGCCTT 671
           ||||||            |||:::   |||:::   :::|||
     954 laArgAlaAlaAspGlyGluGlyAspAspGlyGluArgLysArg...... 968

     672 CAACGGCACTGCAGATAT.....................CGTCAAAAACA 700
            ||||||   |||:::                      |||::::::::
     969 ...ArgHis...ArgHisGlyProProAlaHisAspAspArgGluArgAr 983

     701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
         :|||||||||   ::: |||       |||:::|||         ::::::|||
     983 gHisArgArgArgLys.GluSerGlnGlySerGlyValProMetSerGly 999

     751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTC 800
            :::   :::::::::      :::::::::   ||||||
    1000 ProAsnLeuSerThrThrArgProIleGlnGlnAspLeuGly........ 1013

     801 CACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 850
             ::::::   :::         :::||||||   :::|||  :::|||
    1014 ....ArgGlnAspLeuProLeuAlaGluAspLeuAspAsnMet.LysAsn 1028

     851 TCAAAGACT......................................... 859
         :::|||
    1029 AsnLysLeuAlaThrGlyGluProAlaSerProHisAspSerLeuGlyHi 1045

     860 .......ATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAA 902
                :::|||   :::|||:::   ||||||:::|||   :::|||
    1045 sSerGlyLeuProProSerProAlaLysIleGlyAsnSerThrAsnProG 1062

     903 TGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCC 952
           |||      |||   :::|||
    1062 lyProAlaLeuAlaThrAsnPro........................... 1069

     953 CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATC 1002
                                    :::|||          ||||||||
    1070 ...........................GlnAsn.........AlaAlaSe 1074

    1003 ACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAA 1052
         ||||                             |||:::      :::
    1074 rArg.............................ArgThrProAsnAsnP 1081

    1053 AGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAAT 1102
           |||||||||   :::               |||   :::      |||
    1081 roGlyAsnProSerAsnProGly......ProProLysThrProGluAsn 1095

    1103 ACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1152
         :::      :::|||   |||:::   |||:::|||      |||:::
    1096 SerLeuIleValThrAsnProSer...SerThrGlnProAsnSerAlaLy 1111

    1153 TACGGCAAAGAAAACATCA..................CCTCCTCAACCGT 1184
           ||||||:::|||   :::                  ||||||   :::
    1111 sThrAlaArgLysProGluHisMetAlaValGluIleProPro..AlaCy 1127

    1185 GCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGA 1234
           ||||||   |||         |||::::::      :::   :::|||
    1127 sProProLeuAsnHisThrValValGlnValAsnLysAsnAlaAsnProA 1144

    1235 AGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCAC 1284
                                       :::|||:::   |||::::::
    1144 spPro........................LeuProLysLysGluGluGlu 1152

    1285 GTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGGTAT 1334
            |||   :::   :::   |||                |||
    1153 LysLysGluGluGluGluAlaAsp............ProGlyGluAspGl 1165

    1335 ACCTAAGGCTAAGCCTGTGTTTGATGCG 1362
           ||||||      |||   :::::::::
    1165 yProLysProMetProProTyrSerSer 1174
```